



Universidade de Brasília - UnB
Faculdade UnB Gama - FGA
Engenharia Eletrônica

Estimativa da trajetória do centro de massa corporal utilizando *BlazePose*

Autor: Gustavo Luiz Monteiro de Oliveira
Orientador: Profa. Dra. Claudia Patricia Ochoa Diaz

Brasília, DF
04 de Novembro de 2021



Gustavo Luiz Monteiro de Oliveira

**Estimativa da trajetória do centro de massa corporal
utilizando *BlazePose***

Monografia submetida ao curso de graduação em (Engenharia Eletrônica) da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em (Engenharia Eletrônica).

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Orientador: Profa. Dra. Claudia Patricia Ochoa Diaz

Brasília, DF

04 de Novembro de 2021

Gustavo Luiz Monteiro de Oliveira

Estimativa da trajetória do centro de massa corporal utilizando *BlazePose*/
Gustavo Luiz Monteiro de Oliveira. – Brasília, DF, 04 de Novembro de 2021-
72 p. : il. (algumas color.) ; 30 cm.

Orientador: Profa. Dra. Claudia Patricia Ochoa Diaz

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB
Faculdade UnB Gama - FGA , 04 de Novembro de 2021.

1. centro de massa. 2. engenharia eletrônica. I. Profa. Dra. Claudia Patricia
Ochoa Diaz. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Estima-
tiva da trajetória do centro de massa corporal utilizando *BlazePose*

CDU

Gustavo Luiz Monteiro de Oliveira

Estimativa da trajetória do centro de massa corporal utilizando *BlazePose*

Monografia submetida ao curso de graduação em (Engenharia Eletrônica) da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em (Engenharia Eletrônica).

Trabalho aprovado. Brasília, DF, :

Claudia P. Ochoa D.

**Profa. Dra. Claudia Patricia Ochoa
Diaz**
Orientador

Mariana C. B. Matias

**Profa. Dra. Mariana Costa Bernardes
Matias**
Convidado 1

Roberto de Souza Baptista

Prof. Dr. Roberto de Souza Baptista
Convidado 2

Brasília, DF

04 de Novembro de 2021

Agradecimentos

Agradeço primeiramente à Deus, por iluminar minha trajetória e me conceder saúde, equilíbrio e determinação durante todo o período que durou o desenvolvimento deste trabalho de conclusão de graduação.

Agradeço à minha orientadora Claudia Patricia Ochoa Diaz, pela condução da pesquisa, por todos os ensinamentos, pela constante preocupação com a minha saúde física e mental, pela motivação e dedicação que foram imprescindíveis para o andamento deste trabalho, mesmo durante tempos difíceis impostos pela pandemia.

Sou grato à minha família, em especial aos meus pais, com os quais convivo diariamente. Agradeço a eles pelo amor, confiança e apoio incondicionais proporcionados durante todo o meu período de graduação e, não podendo ser diferente, durante toda a minha vida.

Agradeço ao meu namorado, por estar presente em todos os momentos do meu percurso acadêmico, concedendo apoio e motivação, essenciais para que eu sempre pudesse seguir em frente.

Deixo um agradecimento especial à Universidade de Brasília e à Faculdade do Gama, pela elevada qualidade de ensino e preparação proporcionada a mim e a outros milhares de alunos para um futuro próspero e desafiador, tanto na vida pessoal quanto na vida profissional. Agradeço também à todos os professores e colegas de classe que integraram essa caminhada de maneira tão positiva e enriquecedora.

Resumo

O centro de massa corporal é um parâmetro bastante útil para o estudo do movimento humano. Sua trajetória reflete a interação dos diferentes segmentos corporais na hora de gerar movimento. Existem na literatura vários métodos para estimativa da trajetória do centro de massa corporal e dependendo da natureza dos dados utilizados no cálculo, é necessário um conjunto de equipamentos de alto custo, como câmeras de visão computacional, sistemas de captura de movimento e marcadores corporais, cujo uso é mais apropriado dentro de ambientes laboratoriais. Pensando nas limitações de custo e na complexidade da aplicação dos equipamentos, este trabalho propõe uma abordagem capaz de realizar a estimativa da trajetória do centro de massa automaticamente através da técnica de segmentação corporal, utilizando apenas uma câmera simples e a arquitetura de rede neural *BlazePose*.

Palavras-chaves: centro de massa, *BlazePose*, método de segmentação, captura de movimento

Abstract

The body's center of mass is a very useful parameter for researches regarding human movement. Its trajectory reflects an interaction of different body segments when generating movement. There are several methods in the literature for estimating the trajectory of the body's center of mass and depending on the nature of the data used in the calculation, a set of high-cost equipment is needed, such as computer vision cameras, motion capture systems and body markers, whose use is most suitable within laboratory environments. Considering the cost limitations and the complexity of equipment application, this work proposes an approach capable of automatically estimating the center of mass's trajectory through the body segmentation technique, using only a simple camera and the neural network architecture BlazePose.

Key-words: center of mass, BlazePose, segmental method, motion capture.

Lista de ilustrações

Figura 1 – Comprimentos dos segmentos corporais expressos como fração da altura corporal H . Retirado de (WINTER, 1990).	21
Figura 2 – Localização do centro de massa de um segmento de um corpo em relação à sua massa distribuída. Retirado de (WINTER, 1990)	23
Figura 3 – Um modelo de 14 segmentos com 21 marcadores para estimar o centro de massa 3D do corpo total em experimentos de controle de equilíbrio. Retirado de (WINTER, 1990).	26
Figura 4 – Trajetória vertical do centro de massa corporal ao longo do ciclo de marcha. Retirada de (DIAZ, 2015).	27
Figura 5 – Arquitetura de um neurônio artificial e uma rede neural de multicamadas. Retirado de (BISHOP, 1995).	29
Figura 6 – Modelo <i>Body25</i> com 25 pontos de articulação, desenvolvido a partir do <i>dataset COCO Keypoints</i> . Retirado de (CAO et al., 2018).	31
Figura 7 – Topologia <i>BlazePose</i> das 33 articulações do corpo humano. Retirado de (BAZAREVSKY et al., 2020).	32
Figura 8 – Homem Vitruviano alinhado por meio de dois pontos-chave virtuais previstos através do detector <i>BlazePose</i> , além da caixa delimitadora de rosto. Retirado de (BAZAREVSKY et al., 2020).	33
Figura 9 – Sistema de coordenadas anexado a um corpo rígido. Retirado de (SPONG et al., 2006).	34
Figura 10 – Reprodução de um alinhamento ideal e não-ideal de duas séries temporais. Retirado de (KEOGH E. J.; PAZZANI, 2001).	37
Figura 11 – Matriz de custo das séries temporais. Retirado de (KEOGH E. J.; PAZZANI, 2001).	37
Figura 12 – Visão geral da arquitetura de estimativa de pose humana. Retirado de (BAZAREVSKY et al., 2020).	41
Figura 13 – Saída do Algoritmo 1, mostrando os pontos de articulação de interesse inferidos para 3 posições distintas.	43
Figura 14 – Saída do Algoritmo 2, mostrando os pontos de articulação de interesse conectados entre si para 3 posições distintas.	44
Figura 15 – Saída do Algoritmo 3, mostrando as coordenadas dos pontos de articulação 14, 24 e 25 de acordo com a topologia <i>BlazePose</i>	45
Figura 16 – Centro de massa dos segmentos do corpo em 3 poses distintas.	45
Figura 17 – Centro de massa do corpo todo em 3 poses distintas.	46

Figura 18 – Esboço da vista superior da configuração da sala de captura. A posição das câmeras de vídeo e câmeras de captura de movimento foram organizadas para cobrir um espaço de aproximadamente 3 por 5 metros. Retirado de (GHORBANI et al., 2020).	48
Figura 19 – Vista frontal e lateral do <i>frame</i> de vídeo alinhado, localizações das juntas e pose estimada para um ator feminino e masculino. Retirada de (GHORBANI et al., 2020).	49
Figura 20 – a) início do movimento; b) durante o movimento; c) final do movimento.	51
Figura 21 – Excursão vertical do CoM na marcha com dados do <i>MoCap</i>	52
Figura 22 – Centro de massa corporal na posição a) estática; b) início do ciclo da marcha; c) final do ciclo da marcha. O CoM é representado pela bolinha de cor amarela.	54
Figura 23 – Excursão vertical do CoM para a marcha obtida em a) <i>pixels</i> normalizados e b) milímetros.	55
Figura 24 – Orientação dos eixos no sistema do a) BlazePose; b) MoCap	57
Figura 25 – Excursão vertical do CoM para a marcha obtida em milímetros após transformação homogênea	61
Figura 26 – Excursão vertical do CoM para a marcha obtida em milímetros após transformação homogênea e filtragem	62
Figura 27 – Excursão vertical do CoM nos dois sistemas para a marcha obtida em milímetros.	63
Figura 28 – a) <i>Heatmap</i> com o melhor caminho entre as duas séries pelo DTW; b) Gráfico do DTW aplicado à marcha nos dois sistemas.	64
Figura 29 – Espectro de magnitude da trajetória vertical do CoM nos sistemas a) <i>BlazePose</i> e b) <i>MoCap</i>	66

Lista de tabelas

Tabela 1 – Dados antropométricos dos segmentos corporais. Retirada de (WINTER, 1990).	24
Tabela 2 – Estatística descritiva (Média (Desvio Padrão)) dos 60 atores femininos e 30 masculinos. Retirada de (GHORBANI et al., 2020).	47
Tabela 3 – Visão geral das cinco rodadas de captura, com destaque para a rodada F (Full), selecionada para esse trabalho. Retirada de (GHORBANI et al., 2020).	48
Tabela 4 – Detalhes técnicos dos sistemas de hardware usados para capturar o conjunto de dados <i>MoVi</i> . Retirada de (GHORBANI et al., 2020).	49
Tabela 5 – Frequência e período do coeficiente de maior energia do sinal da trajetória vertical do CoM	65

Lista de abreviaturas e siglas

API	Application Programming Interface
IDE	Integrated Development Environment
CoM	Centro de Massa
CNN	Convolutional Neural Network
RGB	Red Green Blue
2D	Bidimensional
3D	Tridimensional
MOCAP	Motion Capture
COCO	Common Objects in Context
GRF	Ground Reaction Force
ROI	Region of Interest
BML	BioMotionLab
IMU	Inertial Measurement Unit
FPS	Frames por segundo
DTW	Dynamic Time Warping
DFS	Discrete Fourier Series
FFT	Fast Fourier Transform

Sumário

1	INTRODUÇÃO	19
1.1	Considerações iniciais	19
1.2	Definição do problema	19
1.3	Objetivo geral	20
2	FUNDAMENTAÇÃO	21
2.1	Cálculo do centro de massa corporal	21
2.1.1	Métodos convencionais de cálculo do centro de massa	22
2.1.2	Método de segmentação	22
2.1.3	Estudos relacionados	25
2.2	Excursão vertical do CoM durante a marcha	25
2.3	Estimativa da pose humana	26
2.4	Redes Neurais	28
2.5	<i>MediaPipe BlazePose</i>	31
2.5.1	Topologia	31
2.5.2	Arquitetura de detecção de pose	32
2.6	Transformações homogêneas	33
2.7	Filtro de <i>Butterworth</i>	35
2.8	<i>Dynamic Time Warping</i>	36
2.9	Análise espectral do CoM	38
3	DESENVOLVIMENTO	41
3.1	Estimativa da pose humana através de imagens	41
3.1.1	Câmera	41
3.1.2	Inferências dos pontos de articulação	41
3.2	Estimativa do CoM utilizando o método de segmentação no <i>BlazePose</i>	42
3.2.1	Segmentação da estrutura do indivíduo	42
3.2.2	Enumeração dos pontos de articulação com base no modelo <i>BlazePose</i>	43
3.2.3	Cálculo do centro de massa dos segmentos do corpo	45
3.2.4	Cálculo do centro de massa corporal	46
3.3	Base de dados	46
3.4	Excursão vertical do CoM a partir de dados do <i>MoCap</i>	50
3.4.1	Extração dos dados	50
3.4.2	Série temporal com dados do <i>MoCap</i>	51
3.5	Excursão vertical do CoM a partir de dados do <i>BlazePose</i>	52

3.5.1	Série temporal com dados do <i>BlazePose</i>	53
3.5.2	Transformação de coordenadas	56
4	RESULTADOS E DISCUSSÕES	61
4.1	Comparação entre os sistemas	61
4.2	Aplicação do filtro <i>Butterworth</i>	61
4.3	Aplicação do <i>Dynamic Time Warping</i>	62
4.4	Análise espectral da trajetória do CoM	65
5	CONCLUSÃO	67
	REFERÊNCIAS	69

1 Introdução

1.1 Considerações iniciais

O Centro de Massa (CoM) nada mais é do que um ponto imaginário no qual a massa corporal total pode ser considerada como concentrada. As aplicações mais comuns se dão na análise da trajetória do CoM durante a marcha, tendo o deslocamento vertical como um indicador da eficiência da marcha e o horizontal como indicador de equilíbrio e estabilidade.

Existem também as aplicações mais complexas, que envolvem o desenvolvimento de modelos biomecânicos ou ferramentas computacionais capazes de caracterizar a marcha de um grupo de indivíduos, sejam eles saudáveis ou até mesmo os que possuem membros inferiores amputados, através da relação entre movimentos oscilatórios do centro de massa corporal e o centro de pressão do pé, como visto no trabalho de (DIAZ, 2015) e (JUNG C.K.; PARK, 2014).

A importância da estimativa do CoM corporal se faz presente também nos esportes, pois permite classificar o nível de desempenho do atleta e ajuda na descoberta de posições que combinam grandes áreas de apoio e o mínimo de oscilação do CoM, visto que fornecem as melhores condições para dominar o equilíbrio, segundo McKinon (MCKINON et al., 2004).

Na robótica, (MUSCOLO G.; HASHIMOTO, 2011) propõem um método diferente focado na determinação de um algoritmo para encontrar a posição real do CoM de um robô humanoíde bípede usando apenas 2 sensores de força / torque localizados nos pés do robô, um método que permite minimizar o erro entre o robô em si e seu modelo virtual, sem qualquer alteração no sistema de controle ou realização de novos tipos de controle de postura.

1.2 Definição do problema

Muitos pesquisadores, portanto, vêm desenvolvendo métodos para estimativa do CoM baseados em dados adquiridos com equipamentos convencionais como sistemas de captura de movimento (MoCap) e placas de força de reação ao solo ou GRF (*ground reaction force boards*). No entanto, esses sistemas são projetados apenas para ambientes especiais, como laboratórios de pesquisa ou estúdios, e os marcadores são anexados ao indivíduo a ser investigado, o que pode desencadear por exemplo fatores que dificultam a fixação do marcador na pessoa.

Embora tais sistemas sejam confiáveis e precisos, a medição portátil do CoM a

partir de arquiteturas computacionais voltadas para a detecção e classificação de pose humana utilizando imagens e vídeos tem elevado potencial, tendo em vista o avanço da inteligência artificial e aprendizado de máquina, o que permite estabelecer parâmetros de forma automática. Dessa forma, medir o CoM com recursos de hardware limitados e sem custos para o usuário representa uma alternativa promissora no estudo biomecânico deste parâmetro.

1.3 Objetivo geral

Este trabalho propõe um sistema capaz de fornecer a estimativa da trajetória do centro de massa corporal utilizando apenas uma câmera convencional de baixa resolução. Esse sistema será desenvolvido com o auxílio da arquitetura de rede neural para estimativa de pose *BlazePose*.

Dessa forma, é possível pensar em um sistema que tenha como entrada imagens estáticas, vídeos gravados ou em tempo real. Assim, podendo ser utilizado de maneira simples como auxílio em pesquisas sobre o movimento humano.

2 Fundamentação

2.1 Cálculo do centro de massa corporal

A dimensão corporal considerada mais básica é o comprimento dos segmentos entre cada articulação. Estes variam de acordo com a configuração corporal, gênero e origem racial. (DEMPSTER, 1967) resumiu as estimativas de comprimentos dos segmentos e localizações centrais das articulações em relação aos pontos anatômicos. Um conjunto médio de comprimentos segmentares expressados como uma porcentagem da altura corporal é mostrado na figura 1. Segundo (DRILLIS R.; CONTINI, 1966), idealizadores deste parâmetro, estas proporções segmentares podem ser consideradas como uma boa aproximação na ausência de dados mais acurados, medidos diretamente a partir do indivíduo.

O centro de massa (CoM) é um ponto imaginário no qual assume-se que a massa do corpo total está concentrada. A análise da trajetória do CoM no corpo pode fornecer informações importantes sobre equilíbrio e estabilidade corporal do indivíduo. O movimento do CoM durante a caminhada, por exemplo, tem sido frequentemente investigado, pois seu deslocamento vertical é um indicador da eficiência da marcha, de acordo com (BARTONEK A., 2006) e (SAINI et al., 1998), enquanto o movimento horizontal do CoM está relacionado ao equilíbrio e estabilidade, segundo (K. et al., 2009).

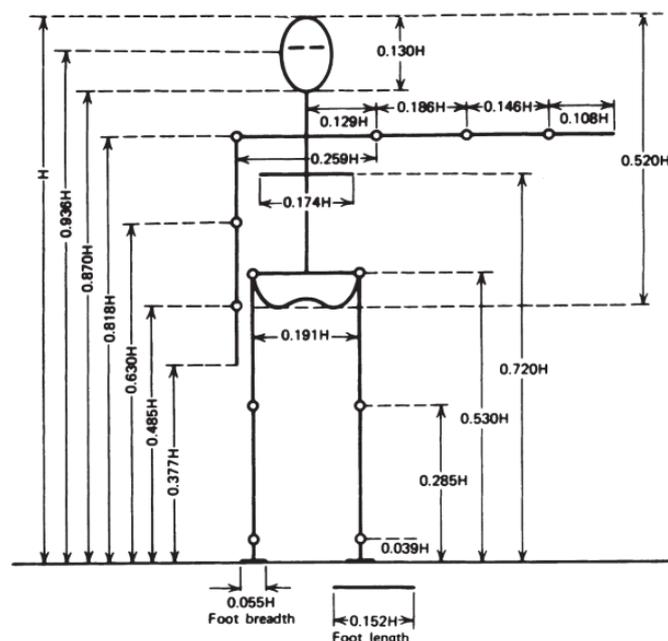


Figura 1 – Comprimentos dos segmentos corporais expressos como fração da altura corporal H . Retirado de (WINTER, 1990).

2.1.1 Métodos convencionais de cálculo do centro de massa

Os métodos comumente utilizados por laboratórios de análise de movimento, segundo (GARD et al., 2004), são:

1. **Método sacral:** método no qual um marcador colocado no sacro, osso grande e triangular localizado na base da coluna vertebral, é considerado uma estimativa do CoM do corpo inteiro.
2. **Método de segmentação:** método em que o CoM de todo o corpo é calculado através das localizações e deslocamentos do CoM de vários segmentos individuais do corpo.
3. **Método cinético:** A trajetória vertical do CoM é obtida através da integração dupla dos dados da força vertical de reação ao solo, medida através de uma plataforma de força de reação ao solo.

(LAFOND D., 2004) e (JAFFREY, 2008) descreveram as abordagens mais utilizadas para a estimativa do CoM e classificaram os métodos de segmentação, baseados em modelos cinemáticos do corpo humano, como os mais adotados em pesquisas, devido à facilidade de implementação. O CoM do corpo é obtido como a soma ponderada da posição do CoM de cada segmento do corpo e o peso da posição CoM de cada segmento é a proporção do segmento para a massa corporal total.

2.1.2 Método de segmentação

O método de segmentação considera a estrutura do corpo como sendo um conjunto de corpos rígidos, identificados por pontos de referência anatômicos. As informações necessárias para esse cálculo são retiradas de tabelas compiladas em estudos antropométricos, os quais compreendem todos os estudos relacionados à medidas físicas do corpo humano que determinam as diferenças entre indivíduos e grupos de indivíduos. Estudos como os de (DEMPSTER, 1967) e (ZATSIORSKY ET AL., 1996) fornecem parâmetros inerciais dos segmentos do corpo e permitem o cálculo de cada segmento. Dessa forma, a localização do CoM de todo o corpo pode ser calculada pelo centroide do sistema multi-segmentar. Por essa razão, essa técnica também é chamada de *segmental kinematic centroid* (SKC) ou, em português, centroide cinemático segmentar.

À medida que a massa corporal total aumenta, também aumenta a massa de cada segmento individual. Portanto, é possível expressar a massa de cada segmento como uma porcentagem da massa total do corpo. A localização do centro de massa é também dada por uma porcentagem do comprimento do segmento desde a extremidade distal ou a proximal. Em estudos realizados com cadáveres, por exemplo, é apontada como simples a

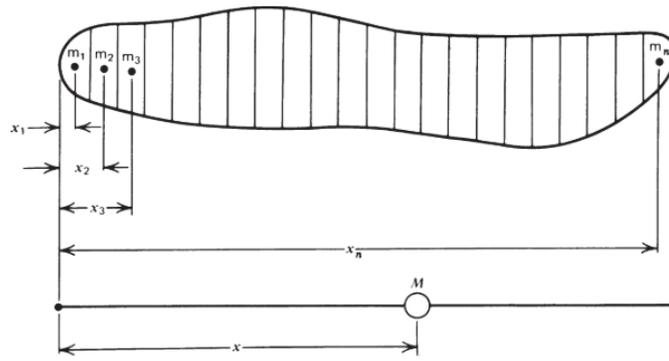


Figura 2 – Localização do centro de massa de um segmento de um corpo em relação à sua massa distribuída. Retirado de (WINTER, 1990)

localização do centro de massa, apenas pela determinação do centro de equilíbrio de cada segmento. Para o cálculo do centro de massa em um corpo humano vivo, é necessário ter o perfil da área da seção transversal do corpo e seu comprimento. A massa total M de cada segmento é expressa pela equação 2.1 abaixo:

$$M = \sum_{i=1}^n m_i \quad (2.1)$$

em que m_i é a massa da i -ésima seção do segmento. Da mesma forma, a figura 2 fornece um perfil hipotético onde o segmento é particionado em n seções, cada uma com a massa indicada.

O centro de massa é tal que deve criar o mesmo momento de força em volta de qualquer ponto ao longo do eixo do segmento assim como a massa distribuída original. Considerando o centro de massa localizado a uma distância x da extremidade esquerda do segmento, tem-se:

$$Mx = \sum_{i=1}^n m_i x_i \quad (2.2)$$

$$x = \frac{1}{M} \sum_{i=1}^n m_i x_i \quad (2.3)$$

Sendo assim, pode-se representar a massa distribuída por uma única massa M localizada a uma distância x de uma extremidade do segmento.

Com cada segmento corporal em movimento, o centro de massa de todo o corpo muda continuamente com o tempo. É, portanto, necessário recalculá-lo após cada intervalo de tempo e isto requer amplo conhecimento das trajetórias do centro de massa de cada segmento do corpo em questão.

TABLE 4.1 Anthropometric Data

Segment	Definition	Segment Weight/Total Body Weight	Center of Mass/ Segment Length		Radius of Gyration/ Segment Length			Density
			Proximal	Distal	C of G	Proximal	Distal	
Hand	Wrist axis/knuckle II middle finger	0.006 M	0.506	0.494 P	0.297	0.587	0.577 M	1.16
Forearm	Elbow axis/ulnar styloid	0.016 M	0.430	0.570 P	0.303	0.526	0.647 M	1.13
Upper arm	Glenohumeral axis/elbow axis	0.028 M	0.436	0.564 P	0.322	0.542	0.645 M	1.07
Forearm and hand	Elbow axis/ulnar styloid	0.022 M	0.682	0.318 P	0.468	0.827	0.565 P	1.14
Total arm	Glenohumeral joint/ulnar styloid	0.050 M	0.530	0.470 P	0.368	0.645	0.596 P	1.11
Foot	Lateral malleolus/head metatarsal II	0.0145 M	0.50	0.50 P	0.475	0.690	0.690 P	1.10
Leg	Femoral condyles/medial malleolus	0.0465 M	0.433	0.567 P	0.302	0.528	0.643 M	1.09
Thigh	Greater trochanter/femoral condyles	0.100 M	0.433	0.567 P	0.323	0.540	0.653 M	1.05
Foot and leg	Femoral condyles/medial malleolus	0.061 M	0.606	0.394 P	0.416	0.735	0.572 P	1.09
Total leg	Greater trochanter/medial malleolus	0.161 M	0.447	0.553 P	0.326	0.560	0.650 P	1.06
Head and neck	C7-T1 and 1st rib/ear canal	0.081 M	1.000	— PC	0.495	0.116	— PC	1.11
Shoulder mass	Sternoclavicular joint/glenohumeral axis	—	0.712	0.288	—	—	—	1.04
Thorax	C7-T1/T12-L1 and diaphragm*	0.216 PC	0.82	0.18	—	—	—	0.92
Abdomen	T12-L1/L4-L5*	0.139 LC	0.44	0.56	—	—	—	—
Pelvis	L4-L5/greater trochanter*	0.142 LC	0.105	0.895	—	—	—	—
Thorax and abdomen	C7-T1/L4-L5*	0.355 LC	0.63	0.37	—	—	—	—
Abdomen and pelvis	T12-L1/greater trochanter*	0.281 PC	0.27	0.73	—	—	—	1.01
Trunk	Greater trochanter/glenohumeral joint*	0.497 M	0.50	0.50	—	—	—	1.03
Trunk head neck	Greater trochanter/glenohumeral joint*	0.578 MC	0.66	0.34 P	0.503	0.830	0.607 M	—
Head, arms, and trunk (HAT)	Greater trochanter/glenohumeral joint*	0.678 MC	0.626	0.374 PC	0.496	0.798	0.621 PC	—
HAT	Greater trochanter/mid rib	0.678	1.142	—	0.903	1.456	—	—

*NOTE: These segments are presented relative to the length between the greater trochanter and the glenohumeral joint.

Source Codes: M, Dempster via Miller and Nelson; *Biomechanics of Sport*, Lea and Febiger, Philadelphia, 1973. P, Dempster via Plagenhoef; *Patterns of Human Motion*, Prentice-Hall, Inc. Englewood Cliffs, NJ, 1971. L, Dempster via Plagenhoef from living subjects; *Patterns of Human Motion*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1971. C, Calculated.

Tabela 1 – Dados antropométricos dos segmentos corporais. Retirada de (WINTER, 1990).

O cálculo do centro de massa corporal é considerado por (WINTER, 1990) um caso especial das equações 2.2 e 2.3. Para um sistema de corpo de n segmentos, o centro de massa nas direções x e y são:

$$x = \frac{m_1x_1 + m_2x_2 + \dots + m_nx_n}{m_1 + m_2 + \dots + m_n} \quad (2.4)$$

$$y = \frac{m_1y_1 + m_2y_2 + \dots + m_ny_n}{m_1 + m_2 + \dots + m_n} \quad (2.5)$$

em que $m_1 + m_2 + \dots + m_n = M$, a massa total do corpo. Sabendo que a massa de cada um dos segmentos é uma porcentagem da massa total M , tem-se que os valores de $m_1 = p_1M$, $m_2 = p_2M$ e assim por diante.

$$x = \frac{p_1Mx_1 + p_2Mx_2 + \dots + p_nMx_n}{M} = p_1x_1 + p_2x_2 + \dots + p_nx_n \quad (2.6)$$

$$y = \frac{p_1My_1 + p_2My_2 + \dots + p_nMy_n}{M} = p_1y_1 + p_2y_2 + \dots + p_ny_n \quad (2.7)$$

Esta equação é mais fácil de usar porque o único requisito é o conhecimento da porcentagem da massa dos segmentos em relação à massa corporal total, representada na equação por p_n , e também presentes na tabela 1 e as coordenadas do centro de massa de cada segmento.

2.1.3 Estudos relacionados

Segundo (CAVAGNA et al., 1977), existe uma relação entre o cálculo do trabalho realizado no CoM e o custo energético para se promover a marcha. Eles explicaram que durante a locomoção em velocidade média, há oscilações tanto da energia potencial quanto da energia cinética do CoM conforme ele sobe e desce, acelera e desacelera, como em um pêndulo invertido.

Já (MACKINNON, 1993), focando em estimativa tridimensional do deslocamento do CoM, utilizou uma estimativa corporal total que compreendia sete segmentos dos membros inferiores e do chamado HAT (*head, arm, trunk*), que compreende os membros superiores cabeça, braços e tronco, para identificar os mecanismos de equilíbrio no plano frontal durante a caminhada nivelada. (JIAN Y.; WINTER, 1993) relataram em seus estudos, uma análise 3D de uma estimativa semelhante de sete segmentos do centro de massa corporal total em conjunto com o centro de pressão durante o início e término da marcha e identificou os mecanismos motores responsáveis por esse movimento tão comum do ser humano.

Um estudo mais recente, realizado por (KAICHI et al., 2019), propõe um método de estimativa do centro de massa 3D, que conta com um conjunto de de múltiplas câmeras e que foi motivado por muitas abordagens baseadas em imagens existentes que analisam o movimento esportivo em jogos reais e nas situações em que é difícil usar dispositivos no corpo dos jogadores. A abordagem proposta por eles foi a primeira tentativa de um processo automatizado de ponta a ponta para estimativa do centro de massa 3D usando apenas entradas de imagem, levando em consideração as propriedades volumétricas do atleta medido.

O protocolo mais completo do centro de massa até hoje foi um modelo de 21 marcadores e 14 segmentos que foi usado para determinar os mecanismos de equilíbrio durante a postura estática ((PATLA F.; WINTER, 1998)). A localização dos marcadores pode ser visto na figura 3, a qual está acompanhada de uma pequena tabela que mostra a localização dos marcadores e a definição de cada um dos 14 segmentos, juntamente com a fração de massa de cada segmento.

2.2 Excursão vertical do CoM durante a marcha

Em um cenário dinâmico, a trajetória do CoM é objetivo fundamental na descrição da marcha humana. Caminhar pode ser definido como uma série de movimentos para cima e para baixo e médio laterais do CoM na direção para frente. Além disso, segundo (GARD et al., 2004), como o CoM representa a localização do peso corporal total, pesquisadores utilizaram a trajetória vertical do CoM durante a caminhada para estimar as mudanças de energia mecânica, para medir a eficiência, para estimar o trabalho, para descreverem

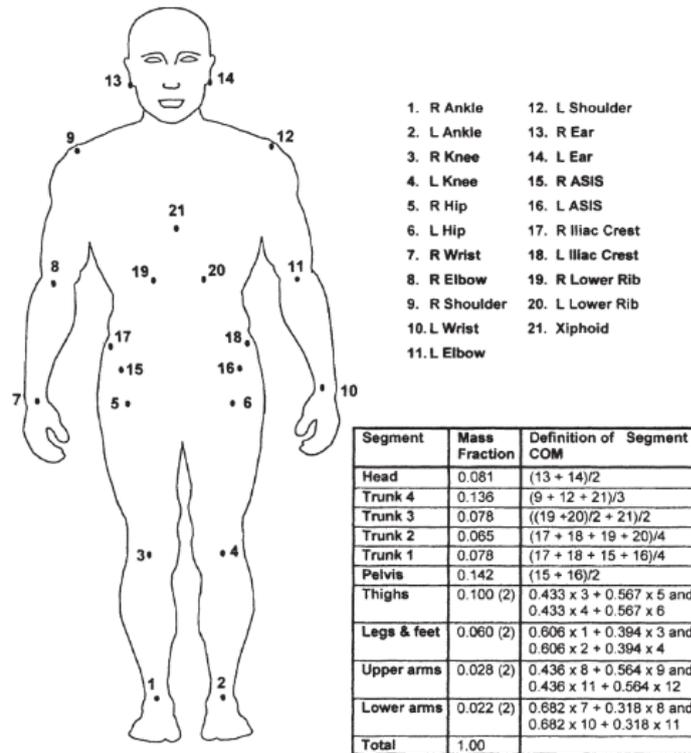


Figura 3 – Um modelo de 14 segmentos com 21 marcadores para estimar o centro de massa 3D do corpo total em experimentos de controle de equilíbrio. Retirado de (WINTER, 1990).

a simetria e para atuar como um indicador da qualidade geral da marcha. Todas essas aplicações dependem da determinação precisa da trajetória vertical do CoM.

A trajetória do CoM descreve um padrão senoidal claro na direção vertical e médio-lateral. O contato inicial é considerado o ponto de referência. A partir dele, a excursão vertical do CoM aumenta à medida que o peso do corpo é projetado na perna de apoio, atingindo seu valor mais alto no apoio médio. Depois disso, esse valor diminui até que a perna de balanço encontre a fase de apoio, ou seja, um novo ciclo de marcha seja iniciado. Em seguida, a sequência é repetida e o peso corporal é agora colocado na perna dianteira. Na Figura 4, que apresenta a excursão vertical do CoM durante um ciclo da marcha, é possível observar que a posição vertical do CoM possui dois pontos mínimo e máximo, que correspondem aos instantes de contato inicial e apoio médio de cada perna. (DIAZ, 2015)

2.3 Estimativa da pose humana

A postura do corpo humano, tradicionalmente, consegue ser reconstruída com acurácia a partir da movimentação capturada através de marcadores ópticos acoplados às partes do corpo, segundo (CHENG S.Y.; TRIVEDI, 2004). Normalmente estes sistemas utilizam múltiplas câmeras para a captura simultânea dos movimentos, entretanto podem

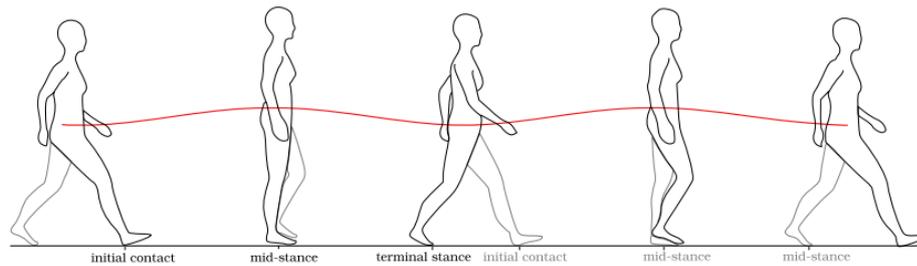


Figura 4 – Trajetória vertical do centro de massa corporal ao longo do ciclo de marcha. Retirada de (DIAZ, 2015).

acabar apresentando desvantagens na implementação por não serem adequados para aplicações não invasivas da vida real, além disso o equipamento é bastante caro e por isso, tem sua praticabilidade bastante limitada, de acordo com (DINH et al., 2014).

Existe uma tendência crescente no número de estudos que focam em métodos que não utilizam os marcadores nos indivíduos. As entradas são as mesmas, as imagens capturadas pelas câmeras e existem diversos tipos de imagens, como a RGB ou imagens em escala de cinza, imagens infravermelhas, imagens com profundidade, dentre outras. Sabendo disso, (SHOTTON et al., 2013) afirma em sua tese que o problema de detecção de pose humana foi mais bem sucedido ao utilizar imagens de profundidade em conjunto com imagens coloridas, por serem comparativamente mais precisas mas esses dispositivos só podem adquirir imagens dentro de um determinado limite de distância, cerca de oito metros, e a grande maioria das imagens da internet são imagens RGB ou em tons de cinza sem informações de profundidade, apenas em algumas ocasiões é apresentada em quatro dimensões (RGB-D), onde as três primeiras correspondem aos canais de cor vermelho, verde e azul, e a quarta dimensão representa a profundidade.

De acordo com (ZHANG X.; GONZÁLEZ, 2016), existem problemas na estimativa da pose humana com características bastante específicas como, por exemplo, os diversos graus de liberdade do corpo humano, o que pode levar a um espaço de soluções de grandes dimensões, e também a estrutura complexa e flexibilidade do ser humano, os quais causam posturas parcialmente oclusas, extremamente difíceis de identificar. Alguns métodos já propostos no passado, como o *visual hull*, conseguiram reduzir esses graus de liberdade justamente com o aumento de câmeras no experimento e limitando o espaço de busca de possíveis configurações corporais às anatomicamente adequadas. No entanto, essas abordagens têm problemas com a identificação precisa da cinemática tridimensional dos segmentos ou utilizam um número limitado de segmentos corporais, como evidenciado no trabalho de (CORAZZA S.; UNDERMANN, 2006).

2.4 Redes Neurais

Recentemente, muitos estudos usando redes neurais para implementar essa estimativa começaram a surgir e se popularizar. A rede neural nada mais é do que um algoritmo ou arquitetura computacional, inspirado na natureza humana e projetado para a resolução de problemas tais como classificação e regressão, a partir da identificação de determinados padrões em dados rotulados.

Segundo (ABRAHAM, 2004), as redes neurais artificiais (ANN) foram desenvolvidas como generalizações de modelos matemáticos de um sistema nervoso. Os elementos básicos de processamento das redes neurais são chamados de neurônios artificiais, ou simplesmente neurônios ou "nós". Em um modelo matemático simplificado do neurônio, os efeitos das sinapses são representados por pesos de conexão que modulam o efeito dos sinais de entrada associados, e a característica não linear exibida pelos neurônios é representada por uma função de transferência. O impulso do neurônio é então calculado como a soma ponderada dos sinais de entrada, transformados pela função de transferência. A capacidade de aprendizagem de um neurônio artificial é alcançada ajustando os pesos de acordo com o algoritmo de aprendizagem escolhido.

A função de transferência da rede neural artificial (ANN), portanto, é formada pelo conjunto de neurônios artificiais, ou seja, o conjunto de "nós" mencionados no parágrafo anterior, em que cada um implementa uma combinação linear de suas entradas \mathbf{x} com pesos \mathbf{W} . Esta combinação é, então, aplicada a uma função de ativação $f(net)$, que representa a não linearidade dos neurônios ou sistema. A saída \mathbf{O} do sistema é expressa por:

$$\mathbf{O} = f(net) = f\left(\sum_{j=1}^n \mathbf{w}_j \mathbf{x}_j\right) \quad (2.8)$$

Na equação 2.8, n é o número de nós da rede e $f(net)$, a função de ativação do nó. A variável net é definida como o produto escalar dos pesos e vetores de entrada, pela equação 2.9 abaixo:

$$net = w^T x = w_1 x_1 + \dots + w_n x_n \quad (2.9)$$

onde T representa a transposição da matriz e, no caso mais simples, o valor de saída O é computado da seguinte maneira:

$$O = f(net) \begin{cases} 1, & \text{se } w^T x \geq \theta \\ 0, & \text{caso contrário} \end{cases} \quad (2.10)$$

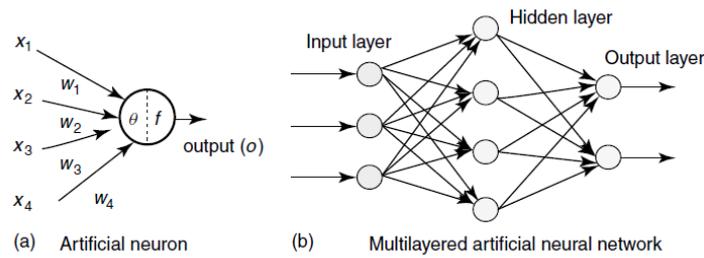


Figura 5 – Arquitetura de um neurônio artificial e uma rede neural de multicamadas. Retirado de (BISHOP, 1995).

onde θ é chamado de nível de limite ou *threshold level* e esse tipo de nó é chamado de unidade de limite linear ou *linear threshold unit*.

A arquitetura básica da ANN, como vista na figura 5 consiste em três tipos de camadas de neurônios: camadas de entrada, ocultas e de saída. Em redes *feed-forward*, o fluxo do sinal é das unidades de entrada para saída, estritamente em uma direção *feed-forward*. O processamento de dados pode se estender por várias camadas de unidades, mas nenhuma conexão de *feedback* está presente, apenas as redes recorrentes contêm conexões de *feedback*. Ao contrário das redes *feed-forward*, as propriedades dinâmicas da rede são importantes. Em alguns casos, os valores de ativação das unidades passam por um processo de "relaxamento" de forma que a rede evolui para um estado estável no qual essas ativações não mudam mais. Em outras aplicações, as mudanças dos valores de ativação dos neurônios de saída são significativas, de modo que o comportamento dinâmico constitui a saída da rede.

Existem várias outras arquiteturas de rede neural, dependendo das propriedades e requisitos da aplicação. A DNN ou rede neural profunda, por exemplo, é uma ANN com múltiplas camadas internas ou ocultas, o que aumenta a quantidade de parâmetros da rede. Já as redes neurais convolucionais (CNN), além dos parâmetros de combinação linear, possui parâmetros de filtros convolucionais aplicados nas primeiras camadas da rede, sendo, portanto, as mais popularizadas no processamento e classificação digitais de imagens. Na CNN, imagens de entrada são processadas por convolução e reduzidas várias vezes para extrair recursos, e as camadas totalmente conectadas consideram todos os esforços integrados. Dessa forma, os erros estimados são propagados de volta (*back propagation*) e os parâmetros de rede são ajustados adequadamente.

Em seu trabalho, (TOSHEV et al., 2014) diz que esses parâmetros são determinados a partir do algoritmo de treinamento (*back propagation*). Este algoritmo consiste em executar a etapa de *feed-forward* para cada dado do conjunto de dados. Através da equação 2.8, o *feed-forward* realiza o cálculo da saída \mathbf{O} em cada nó, com parâmetros iniciados arbitrariamente. Após isso, o valor resultante é comparado com o valor esperado (valor rotulado \mathbf{O}_e), por meio de uma função de erro, conforme a equação 2.11. O objetivo é

minimizar esta função, através de uma otimização do cálculo do gradiente da função custo.

$$E = | O - O_e | \quad (2.11)$$

Além disso, para cada dado rotulado, o *feed-forward* é aplicado e a equação 2.12 é utilizada para atualizar os parâmetros \mathbf{P} das redes neurais em função de cada dado ou conjunto de dados de entrada, não sendo um parâmetro treinado e que deve ser especificado inicialmente. São parâmetros que atuam na sensibilidade da rede, que impede uma espécie de *overflow* em outras camadas de padrões muito específicos aprendidos durante o treinamento, evitando um problema generalizado.

$$P' = P - l_r \nabla E \quad (2.12)$$

em que P' representa os parâmetros atualizados, l_r a taxa de aprendizado e ∇E o gradiente da função custo.

Segundo (HORNİK et al., 1989), redes de multicamadas *feed-forward* com ao menos uma camada escondida (*hidden layer*), são capazes de realizar uma aproximação universal de maneira precisa e satisfatória. Isso significa que uma rede neural satisfaz esse teorema de aproximação universal e consegue resolver qualquer problema relacionado à classificação ou regressão dada uma determinada arquitetura e conjunto de parâmetros, obtidos durante o processo de treinamento dos dados, etapa que demanda bastante memória computacional por, geralmente, exigir uma grande quantidade de dados.

Para a redução da quantidade de dados e também de tempo, é fundamental a aplicação do conceito de transferência de dados durante o treinamento. (PAN S. J.; YANG, 2010) explica em seu trabalho que em muitas aplicações do mundo real, existe uma tarefa de classificação em um determinado domínio de interesse, mas há apenas dados de treinamento suficientes em outro domínio de interesse, onde os últimos dados podem estar em um espaço de recurso diferente ou seguir uma distribuição de dados diferente. Nesses casos, a transferência de dados, se realizada com sucesso, melhoraria muito o desempenho do aprendizado, evitando esforços custosos de rotulagem de dados. Quando se trata, por exemplo, de detecção de rostos em imagens, as camadas iniciais da rede são sensíveis a padrões mais simples como linhas geométricas retas ou circulares e, em seguida, são associados em camadas subsequentes mais complexas de identificação como o formato dos olhos e da boca.

(WU P.; DIETERRICH et al., 2004) desenvolveram um método que também utiliza o conceito de transferência de dados, em que utilizam um conjunto de dados, considerados pequenos e insuficientes, de uma rede neural já treinada sobre imagens de folhas de árvores de bordo e carvalho e, a partir deles, utilizam uma estrutura de dados auxiliares da mesma distribuição dos dados de treinamento e aplicam a um problema de classificação de

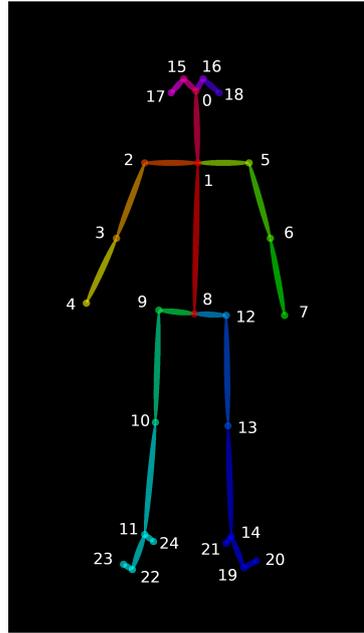


Figura 6 – Modelo *Body25* com 25 pontos de articulação, desenvolvido a partir do *dataset COCO Keypoints*. Retirado de (CAO et al., 2018).

imagens um *kernel* derivado dos formatos que as folhas possuem. Podendo, assim, fornecer um serviço mais amplo de classificação de plantas de outras espécies e informações sobre as mesmas. Ou seja, apenas camadas finais são treinadas novamente a fim de se adequar ao problema e formar entradas para uma rede neural mais específica.

Existem diversas arquiteturas de rede neural com o intuito de realizar a estimativa de pose, como por exemplo: *OpenPose* (CAO et al., 2018), *DensePose* (GÜLER et al., 2018), *PoseFlow* (XIU, 2019), *PoseTrack* (ANDRILUKA, 2018) e *BlazePose* (BAZAREVSKY et al., 2020).

2.5 MediaPipe BlazePose

O *MediaPipe BlazePose* é uma solução que emprega aprendizado de máquina (*Machine Learning*), desenvolvida por (BAZAREVSKY et al., 2020) e tendo a ampla biblioteca *mediapipe* como base, para detecção de pose humana de alta fidelidade, inferindo 33 pontos de referência 2D de um corpo a partir de um único *frame*. Em contraste com o modelo *OpenPose*, baseados na topologia COCO padrão, o *BlazePose* localiza com precisão mais pontos-chave do corpo. Além disso, este método atinge desempenho em tempo real não apenas em computadores e em alguns modelos de celulares.

2.5.1 Topologia

O padrão atual para a estimativa de pose é a topologia COCO (Figura 6), que consiste em 25 pontos de referência no tronco, braços, pernas e rosto. No entanto, os

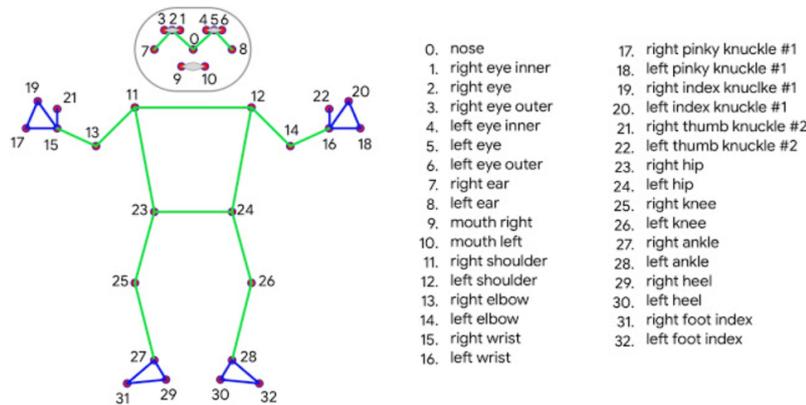


Figura 7 – Topologia *BlazePose* das 33 articulações do corpo humano. Retirado de (BAZAREVSKY *et al.*, 2020).

pontos-chave do COCO localizam-se apenas nos pontos do tornozelo e do punho, o que pode representar uma ausência de informações de escala e orientação para mãos e pés, o que é vital para aplicações práticas como preparação física e dança. A inclusão de mais articulações é essencial na aplicação subsequente de modelos de estimativa de pose específicos de domínio, como aqueles para mãos, rosto ou pés.

Com o *BlazePose*, têm-se, portanto, uma nova topologia que inclui 33 articulações do corpo humano, o que constitui um superconjunto das topologias *BlazeFace* e *BlazePalm*, desenvolvidas por Bazarevsky *et al.* (BAZAREVSKY, 2019) e a COCO. Isso permite ainda determinar a semântica do corpo apenas a partir da predição de pose, que é consistente com os modelos de rosto e mão.

2.5.2 Arquitetura de detecção de pose

Para a estimativa de pose, utiliza-se um arquitetura de detecção em duas etapas. Usando um detector, esta arquitetura primeiramente localiza a região de interesse ou *Region of Interest* (ROI) dentro do *frame*. O detector subsequentemente prevê todos os 33 pontos-chave de pose dessa região de interesse. Para os casos de uso de vídeo, o detector é executado apenas no primeiro *frame* e para os *frames* seguintes, deriva-se a região de interesse dos pontos-chave da pose do *frame* anterior.

Para desempenho em tempo real de todo o *pipeline* de aprendizado de máquina que consiste em modelos de detecção e rastreamento de pose, cada componente deve ser muito rápido, usando apenas alguns milissegundos por *frame*. Para conseguir essa detecção rápida e leve, os desenvolvedores observaram que o sinal mais forte para a rede neural sobre a posição do torso é o rosto da pessoa, devido às suas características de alto contraste e variações comparativamente pequenas na aparência.

Consequentemente, foi treinado um detector de rosto, inspirado no modelo *Blaze-*

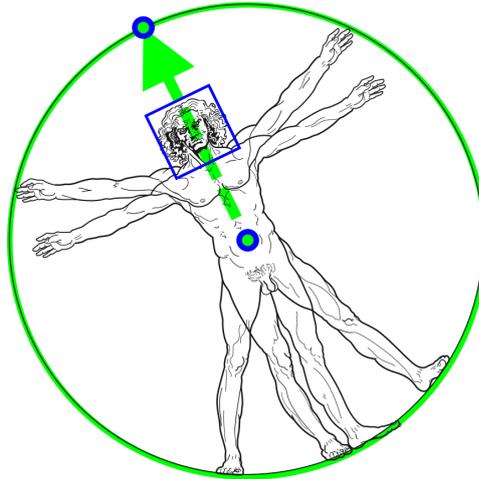


Figura 8 – Homem Vitruviano alinhado por meio de dois pontos-chave virtuais previstos através do detector *BlazePose*, além da caixa delimitadora de rosto. Retirado de (BAZAREVSKY et al., 2020).

Face de (BAZAREVSKY, 2019), como um *proxy* para um detector de pose. Este modelo detecta apenas a localização de uma pessoa dentro do enquadramento e não é usado para identificar indivíduos. Diferentemente dos *pipelines* de rastreamento como o *FaceMesh* e *MediaPipe Hand*, onde deriva-se a região de interesse de pontos-chave previstos, para o rastreamento de pose humana prevê-se explicitamente dois pontos-chave virtuais adicionais que descrevem precisamente o centro do corpo humano, rotação e escala como um círculo. Inspirados pelo homem Vitruviano de Leonardo, como observado na figura 8, eles preveem o ponto médio dos quadris de uma pessoa, o raio de um círculo circunscrevendo a pessoa inteira e o ângulo de inclinação da linha que conecta os pontos médios do ombro e do quadril. Isso resulta em um rastreamento consistente, mesmo para casos muito complicados como imagens de um treino de yoga.

2.6 Transformações homogêneas

Na área de cinemática robótica, existe uma preocupação relacionada com o estabelecimento de vários sistemas de coordenadas para representar as posições e orientações de objetos rígidos, e com as transformações entre esses sistemas de coordenadas. Na verdade, a geometria do espaço tridimensional e dos movimentos rígidos desempenha um papel central em todos os aspectos da manipulação robótica. Estas manipulações envolvem operações de rotação e translação que constituem as transformações homogêneas. (SPONG et al., 2006).

As transformações homogêneas combinam as operações de rotação e translação para construir matrizes de transformação homogêneas, que podem ser usadas para representar simultaneamente a posição e a orientação de um sistema de coordenadas em relação a outro. Além disso, matrizes de transformação homogêneas podem ser usadas para reali-

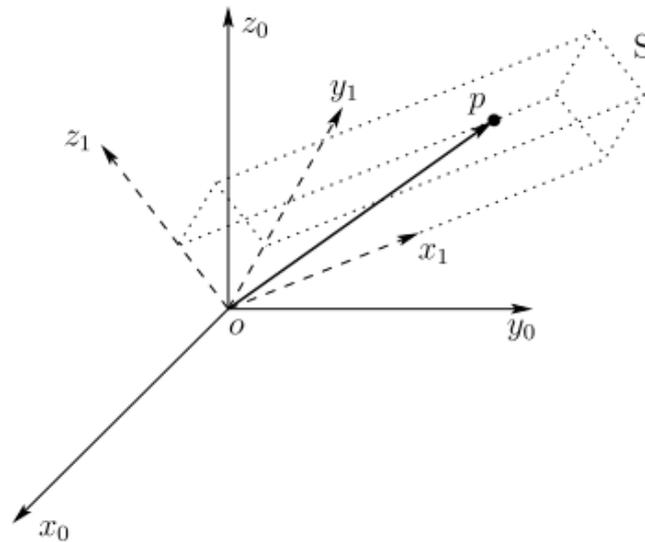


Figura 9 – Sistema de coordenadas anexo a um corpo rígido. Retirado de (SPONG et al., 2006).

zar transformações de coordenadas. Essas transformações permitem a representar vários parâmetros em diferentes sistemas de coordenadas.(SPONG et al., 2006).

A Figura 9, do livro de (SPONG et al., 2006), mostra um objeto rígido S ao qual um sistema de coordenadas $o_1x_1y_1z_1$ está anexo. Dadas as coordenadas p_1 do ponto p (ou seja, dadas as coordenadas de p em relação ao referencial $o_1x_1y_1z_1$), deseja-se determinar as coordenadas de p em relação a um referencial fixo $o_0x_0y_0z_0$. As coordenadas $p_1 = (u, v, w)^T$ satisfazem a seguinte equação:

$$p_1 = ux_1 + vy_1 + wz_1$$

De maneira semelhante, pode-se obter uma expressão para as coordenadas p_0 projetando o ponto p sobre os eixos de coordenadas do sistema $o_0x_0y_0z_0$, fornecendo:

$$p_0 = \begin{bmatrix} p \cdot x_0 \\ p \cdot y_0 \\ p \cdot z_0 \end{bmatrix}$$

Combinando as duas equações, é obtido:

$$p_0 = \begin{bmatrix} x_1 \cdot x_0 & y_1 \cdot x_0 & z_1 \cdot x_0 \\ x_1 \cdot y_0 & y_1 \cdot y_0 & z_1 \cdot y_0 \\ x_1 \cdot z_0 & y_1 \cdot z_0 & z_1 \cdot z_0 \end{bmatrix} \cdot \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

A matriz nesta equação final é meramente a matriz de rotação R , que leva a:

$$p^0 = Rp^1$$

Assim, a matriz de rotação R pode ser usada não apenas para representar a orientação do sistema de coordenadas $o_1x_1y_1z_1$ em relação ao sistema $o_0x_0y_0z_0$, mas também para transformar as coordenadas de um ponto de um sistema para outro. Se um determinado ponto é expresso em relação a $o_1x_1y_1z_1$ pelas coordenadas p_1 , então Rp_1 representa o mesmo ponto expresso em relação ao sistema $o_0x_0y_0z_0$.

Uma transformação homogênea, portanto, nada mais é do que uma representação matricial de um movimento translacional combinado com um rotacional. (SPONG et al., 2006). A transformação homogênea considerada mais geral pode ser escrito como:

$$H = \begin{bmatrix} n_x & s_x & a_x & d_x \\ n_y & s_y & a_y & d_y \\ n_z & s_z & a_z & d_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} & & & d \\ & R & & \\ & & & \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Onde R representa a matriz de rotação e d , a matriz de translação. Nesta equação, $n = (n_x, n_y, n_z)^T$ é um vetor que representa a direção de x_1 no sistema $o_0x_0y_0z_0$, $s = (s_x, s_y, s_z)^T$ representa a direção de y_1 e $a = (a_x, a_y, a_z)^T$ representa a direção de z_1 . O vetor $d = (d_x, d_y, d_z)^T$ representa o vetor da origem o_0 até a origem o_1 expresso no sistema $o_0x_0y_0z_0$.

2.7 Filtro de *Butterworth*

O filtro Butterworth foi caracterizado primeiramente no ano de 1930 em (BUTTERWORTH et al., 1930) e sua principal concepção é ter uma resposta o mais estável possível, com uma sensibilidade uniforme. A sua resposta é plana, não possui *ripple* nem ondulações na banda passante, diferentemente de outras opções de filtros. Além disso, é possível definir diferentes ordens para o filtro de forma a ter uma resposta mais rápida ou mais lenta.

O número de ordens reflete o número de polos do filtro que é a quantidade de elementos reativos como capacitores e indutores necessários para a sua implementação. Filtros de maiores ordens possuem uma resposta mais próxima da ideal, no entanto a construção desse filtro pode ter um alto custo, limitando a aplicação desses filtros à ordens menores em aplicações de baixo custo. (BUTTERWORTH et al., 1930). O ganho $G(\omega)$ de um filtro *Butterworth* passa-baixas de ordem n é fornecido em termos de sua função

de transferência $H(j\omega)$, cuja expressão é apresentada na equação 2.13.

$$G^2\omega = |H(j\omega)|^2 = \frac{G_0^2}{1 + \frac{j\omega}{j\omega_c}^{2n}} \quad (2.13)$$

Onde n é a ordem do filtro, ω_c é a frequência de corte e G_0 o ganho DC na frequência zero. Dessa maneira, à medida que se aumenta o valor de n tendendo ao infinito, o ganho vai se aproximando de uma função retângulo e as frequências abaixo de ω_c ficarão com o ganho G_0 e as frequências acima de ω_c são suprimidas. (BUTTERWORTH et al., 1930).

2.8 Dynamic Time Warping

O *Dynamic Time Warping* (DTW) é um algoritmo que procura distorções entre dois sinais de comportamento semelhante, onde um é o sinal de referência e o outro o alvo da comparação, através do mapeamento dos pontos de maior semelhança. Para isso o sinal alvo é comprimido ou, em outros casos, estendido, buscando a maior similaridade entre as duas sequências temporais. (KEOGH E. J.; PAZZANI, 2001).

O DTW possui aplicações nas sequências temporais de dados de vídeo, áudio e gráficos, de certa forma todos os dados que podem ser transformados em uma sequência linear podem ser analisados com o DTW. Uma aplicação conhecida é o reconhecimento automático de voz, para lidar com diferentes velocidades de fala, visto no trabalho de (SAKOE H.; CHIBA, 1978). Outra aplicação é vista em (BAPTISTA, 2017), que utiliza o DTW para segmentação automática de movimento com base em dados cinemáticos.

A abordagem mais utilizada é determinada por um algoritmo computacional que tem como objetivo encontrar um alinhamento ideal entre duas séries temporais independentes. O algoritmo percorre o eixo y , enquanto distorce o eixo x , resultando em um mesmo ponto sendo conectado a mais de um ponto de outra série. Existe, porém, a possibilidade de o algoritmo não encontrar a correta correspondência entre os pontos e, segundo (KEOGH E. J.; PAZZANI, 2001) isso pode ocorrer devido a um pico ou vale, mais alto ou mais baixo que a referência. A diferença na representação das correspondências ideais (sequências idênticas) e não ideais pode ser vista na Figura 10.

Para a compreensão do algoritmo, um exemplo elementar é detalhado a seguir. Supondo duas séries temporais representadas por A e B , de tamanhos n e m , respectivamente, em que A é representada no eixo x e B no eixo y .

$$A = a_1, a_2, \dots, a_i, \dots, a_n$$

$$B = b_1, b_2, \dots, b_j, \dots, b_m$$

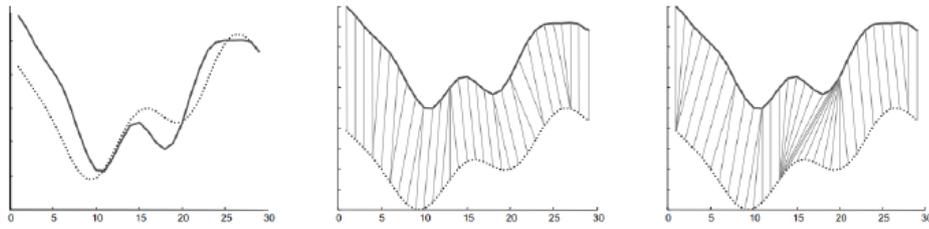


Figura 10 – Reprodução de um alinhamento ideal e não-ideal de duas séries temporais. Retirado de (KEOGH E. J.; PAZZANI, 2001).

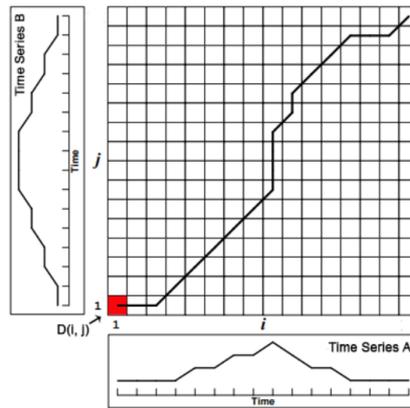


Figura 11 – Matriz de custo das séries temporais. Retirado de (KEOGH E. J.; PAZZANI, 2001).

A matriz $x \times y$, determinada na Figura 11, é construída a partir da distância euclidiana entre os vetores, dada pela equação 2.14.

$$d(i, j) = (a_i - b_j)^2 \quad (2.14)$$

Paralelamente, o algoritmo faz uma comparação entre três valores adjacentes (linha e coluna anterior; linha anterior; coluna anterior) ao valor atual e identifica o menor valor entre eles. O resultado é o elemento $D(i, j)$ armazenando a soma da distância euclidiana com o menor valor, como mostra a equação 2.15.

$$D(i, j) = d(i, j) + \min[D(i - 1, j - 1), D(i - 1, j), D(i, j - 1)] \quad (2.15)$$

Cada elemento (i, j) da matriz condiz com alinhamento entre os pontos (a_i, b_j) . Um caminho W , chamado *Warping Path* representa o conjunto desses elementos que configura o mapeamento entre as séries. Esse caminho pode ser encontrado através da distância cumulativa $\gamma(i, j)$ definida como a distância $d(i, j)$ e o valor mínimo dos elementos adjacentes ao elemento atual. A equação 2.16 descreve as menores distâncias resultantes entre as duas séries, atuando como parâmetro de comparação para correspondência entre todos

os pontos das duas sequências temporais.

$$\gamma(i, j) = d(i, j) + \min[\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)] \quad (2.16)$$

De acordo com (KEOGH E. J.; PAZZANI, 2001), para encontrar esse caminho, as seguintes condições devem ser satisfeitas:

- Fronteira: para garantir que toda a matriz tenha sido percorrida, inicia-se o caminho do primeiro elemento (a_i, b_j) , no canto inferior esquerdo da matriz, e finaliza no canto superior direito (a_n, b_n) .
- Monotonicidade: para garantir que o caminho mapeado seja crescente, ou seja, que o caminho não volta para si mesmo. Além disso, percorrendo-se a diagonal, encontra-se sempre uma correspondência de um ponto do vetor A com um ponto de B .
- Continuidade: deve ser garantido que nenhum elemento da matriz foi ignorado, uma vez que o caminho percorrido não realiza saltos em decorrência do incremento unitário dos índices, percorrendo linhas e/ou colunas.

2.9 Análise espectral do CoM

A característica periódica da trajetória do CoM tem levado pesquisas a usar a Série Discreta de Fourier (DFS) para representar matematicamente o desempenho deste parâmetro durante um ciclo de marcha. De acordo com a formulação DFS, uma sequência periódica discreta $x[n]$ com um período de N amostras pode ser representada como uma soma de funções senoidais cujas frequências são múltiplas da frequência fundamental $\omega_0 = \frac{2\pi}{N}$. (DIAZ, 2015).

Portanto, a Série Discreta de Fourier (DFS) da sequência $x[n]$ é dada pela seguinte expressão:

$$x[n] = \sum_{k=1}^N a_k e^{jk(\frac{2\pi}{N})n} \quad (2.17)$$

Da mesma maneira, os coeficientes DFS podem ser escritos como:

$$a_k = \sum_{n=1}^N x[n] e^{-jk(\frac{2\pi}{N})n} \quad (2.18)$$

Como a sequência $x[n]$ tem um período de N amostras, seu espectro só pode ter componentes que estão localizados em frequências múltiplas de ω_0 . Isso significa que os coeficientes DFS aparecerão em frequências definidas como:

$$\omega_k = k\omega_0 \quad \text{com } k = 1, 2, 3, \dots, N \quad (2.19)$$

Para definir quantos coeficientes DFS são necessários para ter uma boa aproximação de $x[n]$, o cálculo da energia do sinal é um critério útil.([DIAZ, 2015](#)). A energia no domínio da frequência pode ser calculada em termos dos coeficientes DFS, como:

$$E_t = \sum_{k=1}^N |a_k|^2 \quad (2.20)$$

em que $|a_k|^2$ é conhecida como a potência média do k -ésimo componente do DFS.

3 Desenvolvimento

Este capítulo descreve as etapas realizadas para a estimativa e análise da trajetória do centro de massa do corpo de um indivíduo durante a marcha, calculado através do método de segmentação e utilizando a arquitetura de rede neural *BlazePose*, bem como a validação desta ferramenta por meio de comparação com dados de posição de um sistema de captura de movimento (*MoCap*). A plataforma utilizada para o desenvolvimento e descrição do código foi a *Jupyter Notebook* e a linguagem *Python*.

3.1 Estimativa da pose humana através de imagens

3.1.1 Câmera

Elemento protagonista na primeira etapa, a câmera é um instrumento de medição e único sensor do trabalho. A câmera fornece a imagem estática e o vídeo, os quais podem ser vistos como matrizes $I_{H \times W \times 3 \times N}$, em que H e W representam a altura e largura do *frame*, respectivamente, e N o número de *frames*. A câmera utilizada para o sistema foi a câmera do celular iPhone 11, com resolução de 1080p e taxa de amostragem de 30 fps.

3.1.2 Inferências dos pontos de articulação

Para a parte de implementação em *Python*, dois principais recursos utilizados foram as bibliotecas *OpenCV* e *mediapipe*, a primeira usada para a etapa de processamento da imagem e a última, para a estimativa de pose.

Este passo abrange o processamento de dados vindos da imagem captada pelo sensor para a estimativa de pose da pessoa filmada. Portanto, utiliza-se a API do *BlazePose* e como entrada da rede, o vídeo $I_{H \times W \times 3 \times N}$. A saída esperada é, portanto, o vídeo com a estimativa dos pontos de articulação $O_{P \times J \times 2 \times N}$, em que P é o número de pessoas encon-

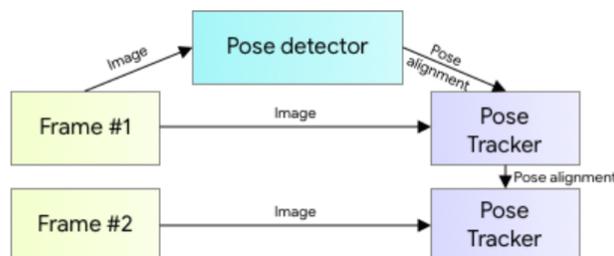


Figura 12 – Visão geral da arquitetura de estimativa de pose humana. Retirado de (BAZAREVSKY et al., 2020).

tradas, J o número de articulações, 2 representa a dimensão da coordenada em *pixels* de cada ponto e N o número de *frames*. A API utilizará o modelo de articulação *BlazePose*, como já discutido, baseado em modelos já conhecidos como COCO, MPII e BODY-25, obtidos pelo treinamento da CNN com diferentes e extensos *datasets*.

O primeiro teste da rede foi feito com um vídeo simples de um indivíduo realizando algumas atividades de treino físico, envolvendo a movimentação de membros superiores e inferiores do seu corpo.

O *BlazePose* consiste em um detector de pose, seguido por uma rede de rastreadores de pose. No caso de vídeos, o rastreador primeiramente prevê as coordenadas dos pontos-chave (articulações), a presença da pessoa no *frame* atual e a região de interesse para este *frame*. Quando o rastreador indica que não há nenhum humano presente, a rede de detectores é reexecutada no próximo *frame*. Esse modelo de detecção da pose é ilustrado na Figura 12.

O Algoritmo 1 é capaz de definir e mostrar os pontos de interesse, cujo conjunto compõe a estimativa completa da pose, como mostrado na Figura 13.

Algoritmo 1: Ler o vídeo e retornar a estimativa de pose.

```

cap = cv2.VideoCapture('PoseVideo.mp4')
while True do
    success, img = cap.read()
    imgRGB = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
    results = pose.process(imgRGB)
    print(results.pose_landmarks)
    if results.pose_landmarks then
        | mpDraw.draw_landmarks(img, results.pose_landmarks)
    end
end
end

```

3.2 Estimativa do CoM utilizando o método de segmentação no *BlazePose*

3.2.1 Segmentação da estrutura do indivíduo

Após a inferência dos pontos de articulação na pessoa, inicia-se o passo de interligação entre esses pontos, o que irá retornar a estrutura completa do corpo do indivíduo em questão, formado por diferentes segmentos, que nada mais são do que as linhas que

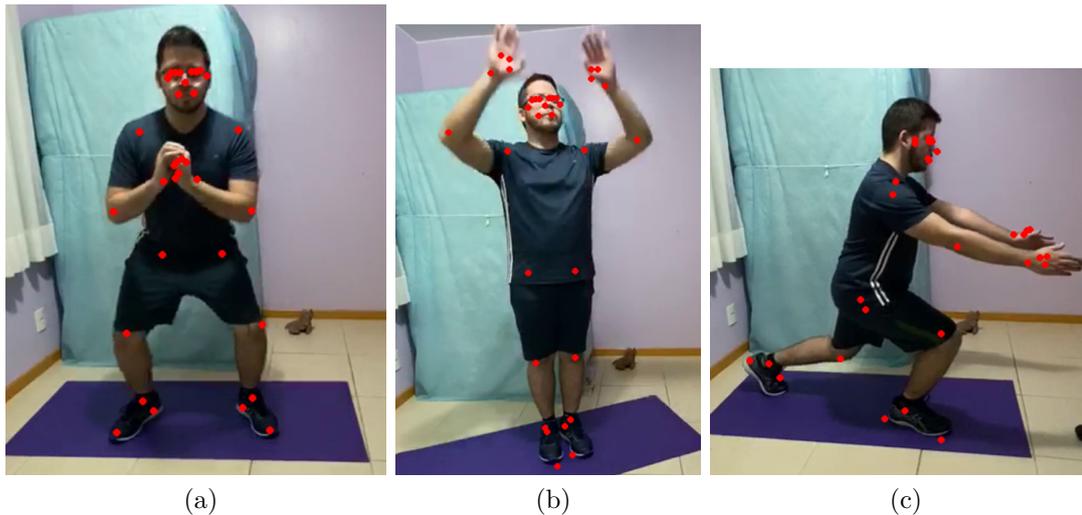


Figura 13 – Saída do Algoritmo 1, mostrando os pontos de articulação de interesse inferidos para 3 posições distintas.

conectam uma coordenada à outra, como apresentado na Figura 14.

Algoritmo 2: Interligar os pontos das articulações estimados.

```

cap = cv2.VideoCapture('PoseVideo.mp4')
while True do
    success, img = cap.read()
    imgRGB = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
    results = pose.process(imgRGB)
    print(results.pose_landmarks)
    if results.pose_landmarks then
        mpDraw.draw_landmarks(img, results.pose_landmarks,
            mpPose.POSE_CONNECTIONS)
    end
end
end

```

A única mudança no Algoritmo 2 para esta etapa, foi a inclusão do desenho das conexões entre os pontos através de uma função da biblioteca mpPose, destacada em negrito.

3.2.2 Enumeração dos pontos de articulação com base no modelo *BlazePose*

Após a etapa de segmentação, para se conhecer as coordenadas dos 33 pontos de articulação da topologia *BlazePose* que se conectam e formam o segmento no plano da imagem, foi necessário criar uma estrutura de repetição do tipo *for* para que haja a extração de todas as coordenadas dos pontos em tempo real para cada *frame* da imagem. O Algoritmo 3 traz, portanto, o laço de repetição da extração das coordenadas $[x,y]$ dos pontos de articulação normalizadas para $[0,0, 1,0]$ pela largura e altura da imagem,

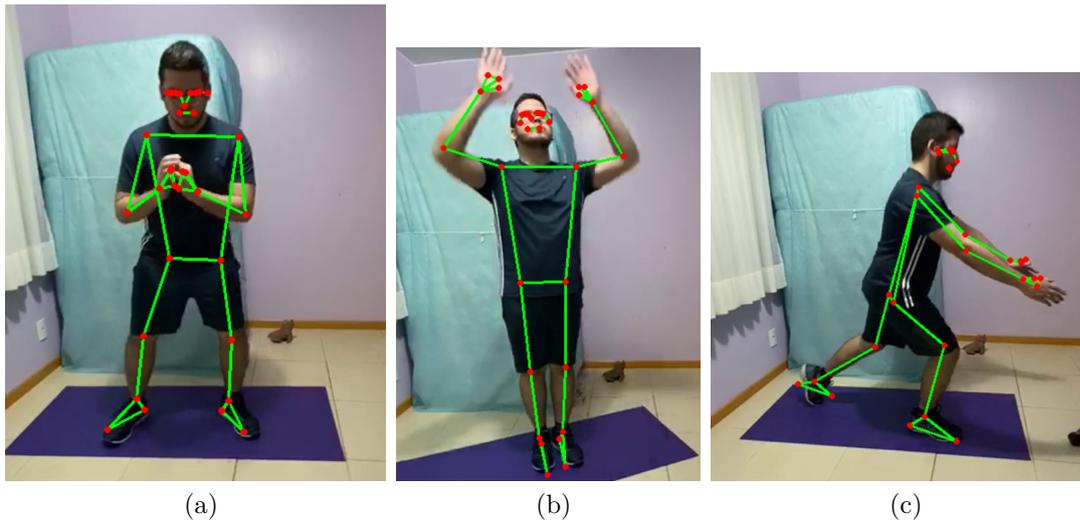


Figura 14 – Saída do Algoritmo 2, mostrando os pontos de articulação de interesse conectados entre si para 3 posições distintas.

respectivamente.

Algoritmo 3: Enumeração dos pontos de articulação e suas coordenadas.

```

cap = cv2.VideoCapture('PoseVideo.mp4')
while True do
    success, img = cap.read()
    imgRGB = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
    results = pose.process(imgRGB)
    print(results.pose_landmarks)
    if results.pose_landmarks then
        mpDraw.draw_landmarks(img, results.pose_landmarks,
            mpPose.POSE_CONNECTIONS)
        for id, lm in enumerate(results.pose_landmarks.landmark) do
            h, w, c = img.shape
            print(id, lm)
            cx, cy = int(lm.x * w), int(lm.y * h)
            cv2.circle(img, (cx, cy), 5, (255, 0, 0), cv2.FILLED)
        end
    end
end
end

```

Percebe-se, portanto, que as coordenadas de cada ponto são enumeradas de 0 a 32, totalizando os 33 pontos do modelo *BlazePose*. Dessa forma, basta realizar uma análise em paralelo entre esta topologia e a imagem de saída com os pontos de articulação em destaque para saber a quais membros corporais se referem àquelas coordenadas. Um exemplo do resultado para a imagem utilizada pode ser visto na Figura 15, em que os números 14, 24 e 25 referem-se às articulações do cotovelo esquerdo, quadril esquerdo e

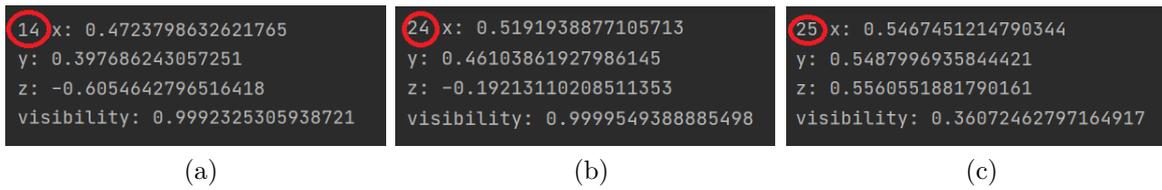


Figura 15 – Saída do Algoritmo 3, mostrando as coordenadas dos pontos de articulação 14, 24 e 25 de acordo com a topologia *BlazePose*.

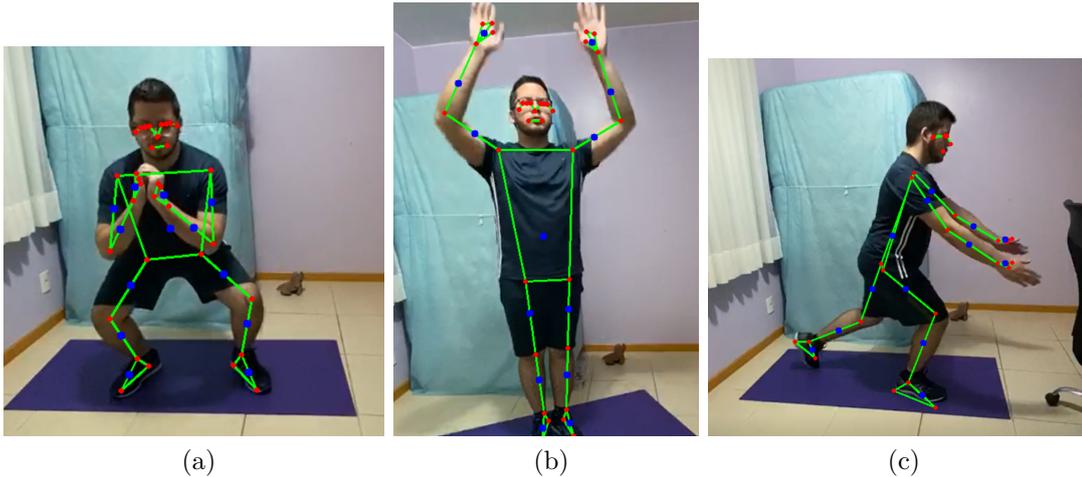


Figura 16 – Centro de massa dos segmentos do corpo em 3 poses distintas.

joelho direito, respectivamente.

3.2.3 Cálculo do centro de massa dos segmentos do corpo

Após a obtenção das coordenadas dos pontos das articulações, foi desenvolvido o algoritmo para extrair os CoM dos segmentos do corpo necessários para o cálculo do CoM corporal. Para esta etapa, foi realizado uma análise paralela entre o modelo de corpo humano *BlazePose*, da Figura 7, e a Tabela 1 de dados antropométricos dos segmentos corporais. Dessa forma, foi possível observar, a partir da segunda coluna da tabela, a definição de cada segmento, ou seja, a exata localização das articulações corporais que abrangem aquele segmento e, simultaneamente, correlacioná-los aos segmentos apresentados no modelo em questão.

Seguindo a tabela de dados antropométricos, as posições relativas do CoM para cada segmento são dadas como uma porcentagem de seu comprimento a partir da extremidade mais próxima ao CoM, também conhecida como extremidade proximal. Sendo assim, as posições das articulações detectadas foram usadas para calcular o comprimento de cada segmento e, em seguida, multiplicá-los pelas suas determinadas posições relativas para, finalmente, chegar ao valor estimado do CoM do segmento e plotá-lo na imagem de saída.

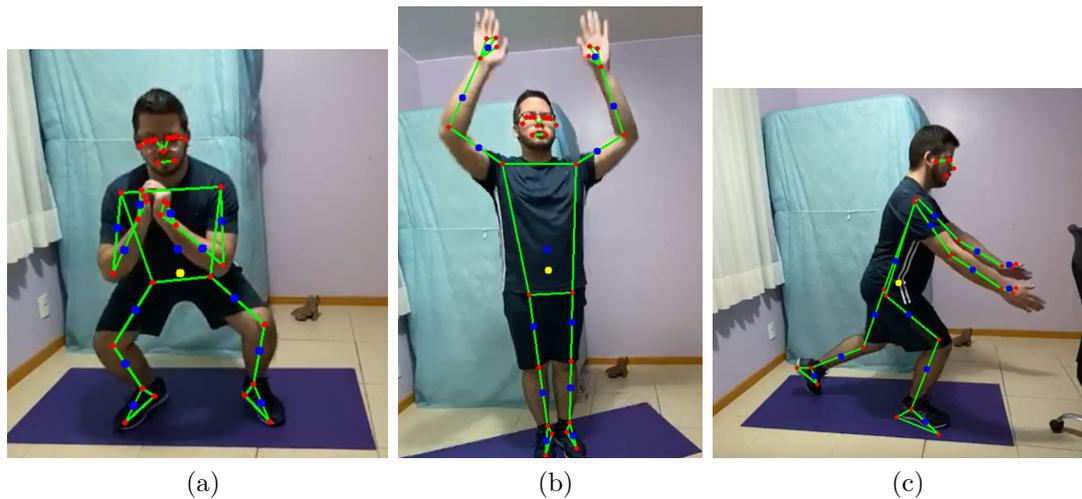


Figura 17 – Centro de massa do corpo todo em 3 poses distintas.

Os CoM dos segmentos do corpo em 3 poses distintas pode ser visto na Figura 16, representados pelos pontos circulares plotados na cor azul. Portanto, é possível constatar que a localização anatômica dos CoM dos segmentos nas imagens é consistente com os dados das posições relativas tomados como referência da Tabela 1, em relação às extremidades proximais de cada segmento.

3.2.4 Cálculo do centro de massa corporal

O CoM corporal em 3 poses distintas pode ser visto na Figura 17, representados pelos pontos circulares plotados na cor amarela, calculados a partir das equações 2.6 e 2.7. Como discutido na seção 2.1.1, estima-se que a localização anatômica do CoM corporal fica em uma região próxima ao sacro, uma estrutura óssea localizada na base das vértebras lombares e conectada à pelve. Nas imagens, é possível observar que a estimativa do centro de massa corporal mostra-se razoável, considerando que aproximadamente 79% dos segmentos corporais compõem este cálculo do CoM corporal, como visto também na Tabela 1.

3.3 Base de dados

Seguindo o propósito de validação da ferramenta *BlazePose* para o cálculo do centro de massa corporal e estimativa de sua trajetória no tempo para determinado movimento, foram utilizados dados fornecidos pela base de dados *MoVi*, do trabalho *MoVi: A Large Multipurpose Motion and Video Dataset* de (GHORBANI et al., 2020).

MoVi é um base de dados de movimento projetado para uma variedade de desafios, como estimativa de pose humana, reconhecimento de ação, modelagem de movimento, análise de marcha e reconstrução da forma corporal. Essa base de dados está associado

	Females	Males
Height (m)	1.65 (0.08)	1.75 (0.06)
Weight (kg)	60.35 (8.03)	72.3 (10.98)
BMI (kg/m ²)	22.16 (3.02)	23.6 (3.24)
Age (y)	20.47 (3.59)	23.6 (3.61)

Tabela 2 – Estatística descritiva (Média (Desvio Padrão)) dos 60 atores femininos e 30 masculinos. Retirada de (GHORBANI et al., 2020).

ao laboratório *BioMotionLab* que faz parte da York University de Toronto, no Canadá. O acesso ao *BML-MoVi* está disponível em um repositório canadense para fins de pesquisa.

Ele contém 60 atores femininos e 30 masculinos, sendo 5 pessoas canhotas, realizando uma coleção de movimentos que incluíam ações cotidianas e movimentos esportivos: (1) Andar, (2) Correr, (3) Correr no lugar, (4) Galope lateral, (5) Rastejar, (6) Salto vertical, (7) Macacos de salto, (8) Chutar, (9) Alongar, (10) Cruzar os braços, (11) Sentar-se em uma cadeira, (11) Cruzar as pernas enquanto está sentado, (13) Apontar, (14) Bater palmas, (15) Coçar a cabeça, (16) Jogando e pegando, (17) Acenando, (18) Fingindo tirar uma foto, (19) Fingindo falar ao telefone, (20) Fingindo checar o relógio. Em cada uma das cinco sequências, os atores realizaram adicionalmente um movimento escolhido por eles (21). É importante ressaltar que antes de iniciar cada movimento, o sujeito fica em uma pose estática de repouso. As estatísticas descritivas de todos os participantes são mostradas na Tabela 2.

Em cinco rodadas de gravações, onde cada uma das rodadas contém os 20 movimentos, os mesmos atores e movimentos foram gravados usando diferentes sistemas de hardware, incluindo um sistema de captura de movimento óptico (*MoCap*), câmeras de vídeo e unidades de medição inercial (*IMU*), dispostos no estúdio de gravação da maneira apresentada na Figura 18. Para algumas rodadas de captura os atores foram gravados vestindo roupas convencionais, para as outras rodadas eles usaram poucas peças de roupas. O procedimento experimental foi aprovado pelo comitê de ética da Queen’s University, Kingston, e foi realizado de acordo com a Declaração de Helsinque.

A sequência de captura de dados intitulada “F” foi capturada usando o layout de 67 marcadores *MoCap* sugerido em (LOPER et al., 2014). Os atores usavam o mínimo de peças de roupas possível e justas para minimizar o movimento do marcador em relação ao corpo. Os marcadores foram fixados na pele e nas roupas dos atores com fita dupla-face. O sistema *MoCap* foi sincronizado com duas câmeras de vídeo capturando as ações de diferentes pontos de vista (frontal e lateral). Essas duas câmeras foram calibradas computando-se a translação e rotação das câmeras em relação ao sistema de coordenadas

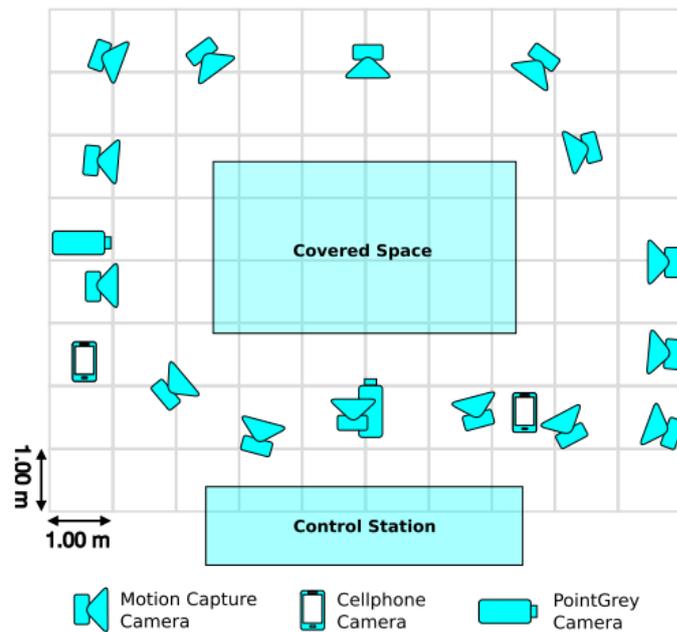


Figura 18 – Esboço da vista superior da configuração da sala de captura. A posição das câmeras de vídeo e câmeras de captura de movimento foram organizadas para cobrir um espaço de aproximadamente 3 por 5 metros. Retirado de (GHORBANI et al., 2020).

Round	F	S1	S2	I1	I2
MoCap marker set	67	12	12	–	–
Video capture	yes	yes	yes	yes	yes
IMU	no	yes	yes	yes	yes
A-pose between motions	yes	yes	no	yes	no
Actor clothing	minimal	normal clothing	normal clothing	normal clothing	normal clothing
Length (min per person)	~ 2.7	~ 2.7	~ 1.7	~ 2.7	~ 1.7

Tabela 3 – Visão geral das cinco rodadas de captura, com destaque para a rodada F (Full), selecionada para esse trabalho. Retirada de (GHORBANI et al., 2020).

do sistema *MoCap*.

A motivação para a escolha desta rodada de captura para obtenção de dados para a trajetória do CoM na marcha foi devido a quantidade adequada de marcadores no corpo, como mostra a Tabela 3. Imagens de exemplo de um ator feminino e masculino em pose de repouso na rodada F são mostradas na Figura 19.

O sistema de captura ótica de movimento comercial utilizado no desenvolvimento da base de dados foi da Qualisys com quinze câmeras de 1.3 Mega Pixels que fornecem a localização 3D de marcadores reflexivos passivos com uma taxa de 120 *frames* por segundo. Os dados de vídeo foram coletados usando dois tipos diferentes de câmeras, câmeras de smartphones e câmeras de visão computacional, chamadas de *PointGrey Cameras*, as quais podem ser visualizadas na Figura 18.

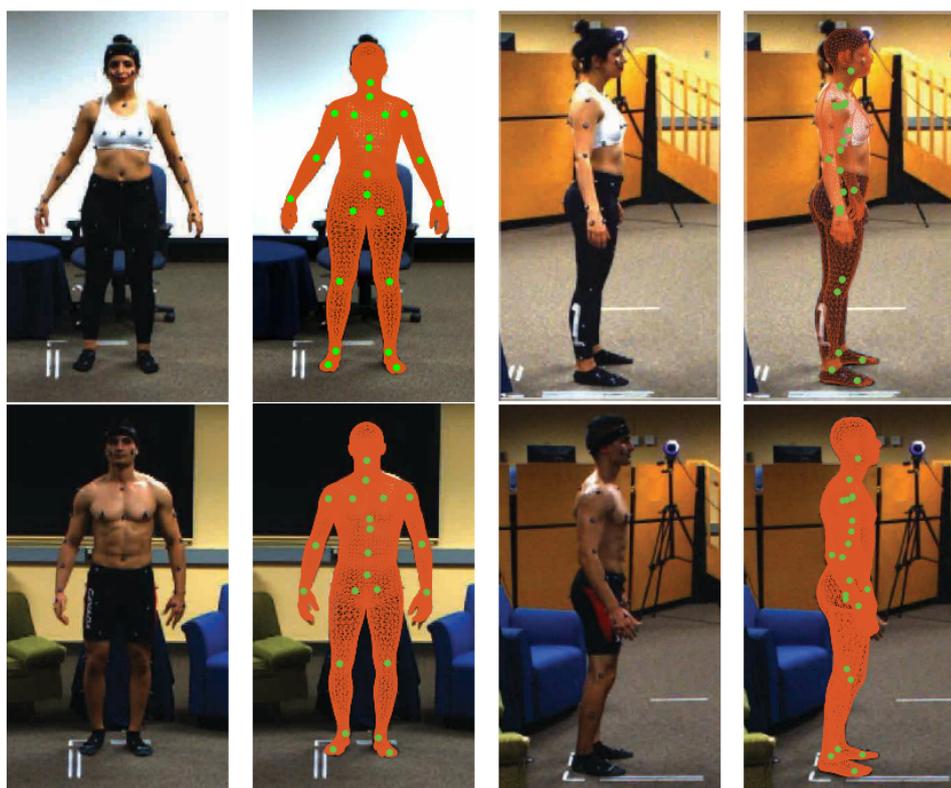


Figura 19 – Vista frontal e lateral do *frame* de vídeo alinhado, localizações das juntas e pose estimada para um ator feminino e masculino. Retirada de (GHORBANI et al., 2020).

MoCap System	Brand and model	Qualisys Oqus 300 and 310
	Number of cameras	15
	Resolution	1.3 megapixel
	Frame rate	120 Hz
	Synchronized	Yes
Video Capture Systems 1	Brand and model	FLIR, Grasshopper
	Number of cameras	2
	Cameras synchronized	Yes
	Cameras calibrated	Yes
	Resolution	800 x 600 pixels, 72 dpi, 24 bit depth
	Sensor	Sony ICX285, CCD
	Frame rate	30 Hz
	File type	JPEG frames and AVI video files

Tabela 4 – Detalhes técnicos dos sistemas de hardware usados para capturar o conjunto de dados *MoVi*. Retirada de (GHORBANI et al., 2020).

Para as duas câmeras de visão computacional, usaram-se câmeras do modelo *Grasshoppers* da empresa *FLIR Inc.* com sensores *CCD Sony ICX285* de 800×600 pixels. As gravações das câmeras *FLIR* são sincronizadas com as câmeras *MoCap* com 30 fps (alinhadas a cada quarto *frame* do sistema *MoCap*), como exposto detalhadamente na Tabela 4.

Com o objetivo de fornecer um movimento 3D preciso *frame a frame* sobreposto ao vídeo, o sistema de captura de movimento é sincronizado com as câmeras no *frame* e, em seguida, calibrado para a mesma coordenada. Segundo os desenvolvedores da base de

dados, as câmeras de vídeo são acionadas através do sinal de sincronização fornecido pelo sistema *MoCap* e devido aos limites de taxa de *frames* em câmeras de vídeo, a frequência de sincronização foi dividida por 4, o que reduziu a taxa de *frames* de captura de vídeo para 30 fps.

3.4 Excursão vertical do CoM a partir de dados do *MoCap*

Primeiramente, para acessar os dados de *MoCap* da base de dados *MoVi*, foi executado um código atribuído por (GHORBANI et al., 2020), juntamente com um tutorial no *Jupyter Notebook* para a leitura dos arquivos *.mat* (*Microsoft Access Table*) em *Python*.

Esses arquivos contém os dados *MoCap* processados por V3D ou *Visual 3D*, um software de análise biomecânica avançada para dados de captura de movimento 3D. O V3D fornece uma estimativa mais precisa da localização da articulação, que apesar de ser considerado pelos desenvolvedores do *dataset* como menos robusto para evitar oclusões, é preferível na análise da marcha.

Todos os arquivos apresentam-se compactados e contém dados dos 90 sujeitos em todas as rodadas. A partir da abertura dos dados da rodada F, foram selecionados, portanto, as informações de interesse para esse trabalho, que serão descritas na subseção 3.4.1.

3.4.1 Extração dos dados

Como o movimento protagonista desse trabalho é a marcha, foi necessário separar os dados desse movimento para realizar a análise. Para tanto, utilizou-se os dados do vídeo como forma de realizar a segmentação dos dados de *MoCap*, também disponibilizado de forma sincronizada com o vídeo. Ainda, optou-se pelo foco apenas na excursão vertical do CoM, uma vez que do ponto de vista biomecânico, este deslocamento possui os maiores valores e têm destaque no estudo de parâmetros de avaliação da marcha e equilíbrio.

Mesmo com a taxa de amostragem dos vídeos de 30 Hz e a taxa de amostragem do *MoCap* de 120 Hz, foi possível converter o tempo do vídeo no tempo dos dados de *MoCap* e separar os dados de *MoCap* com informações somente referentes à marcha. Os dados de *MoCap* coletados foram os dados de deslocamento, em *mm* (milímetros), para os eixos *x*, *y* e *z* dos marcadores localizados nas articulações imprescindíveis para o cálculo do deslocamento vertical do CoM na marcha.

Os dados de deslocamento foram adotados como a entrada da rotina, para cada *frame* correspondente ao movimento de marcha, que é sistematicamente a equação 2.6 e a 2.7, em que M é a massa total do corpo e p_n , a porcentagem da massa dos segmentos em relação à massa do corpo todo, fornecidas pela Tabela 1. O vídeo utilizado para este

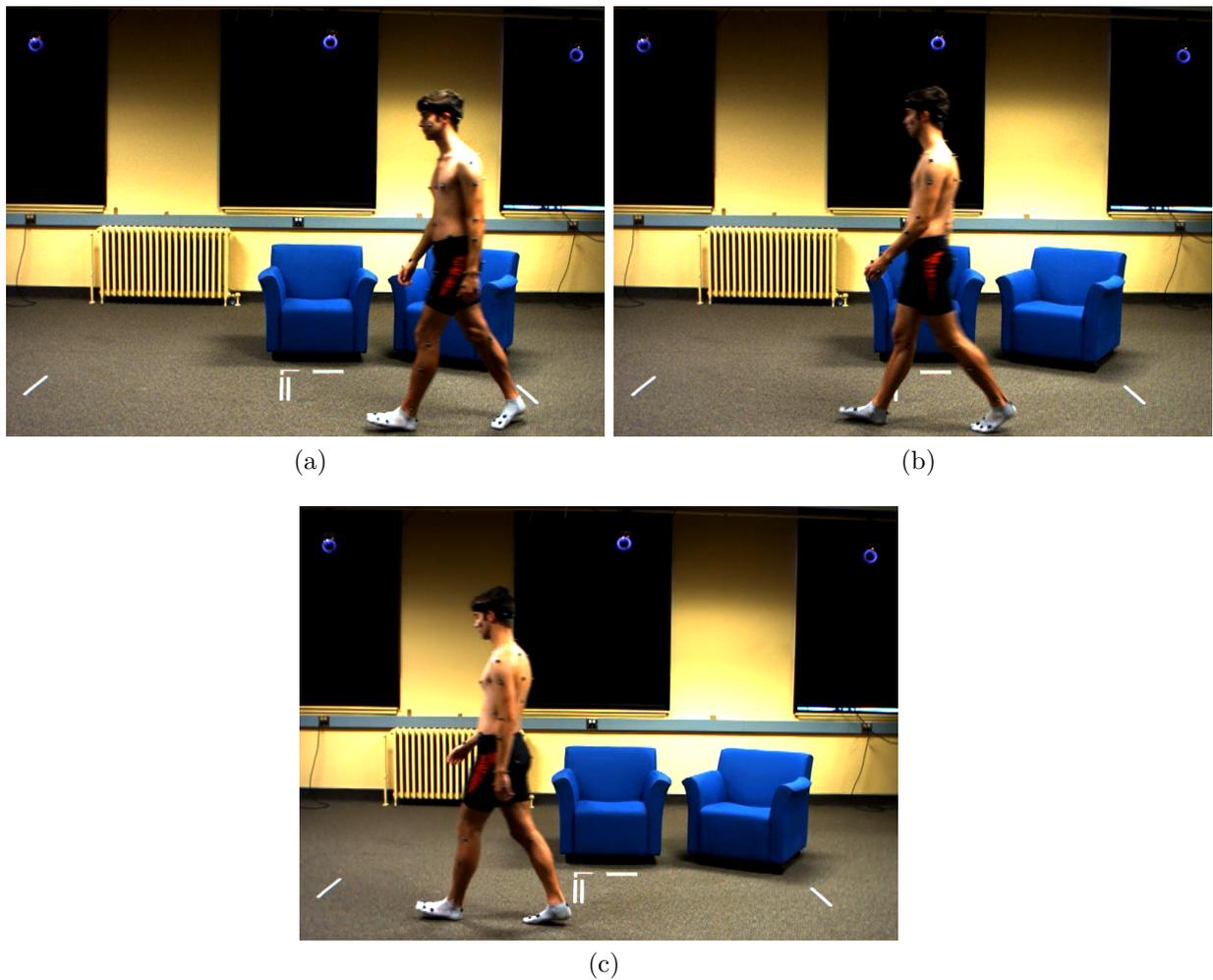


Figura 20 – a) início do movimento; b) durante o movimento; c) final do movimento.

procedimento de segmentação de dados foi o do voluntário de número 15, um indivíduo de 21 anos do sexo masculino, destro, altura de 1.81 *m* e pesando 72 *kg*, selecionado aleatoriamente. No vídeo, é possível ver o ciclo da marcha do sujeito começando quando o pé esquerdo toca o chão, na extremidade direita da tela e terminando mais à esquerda, quando o mesmo pé encontra o chão novamente, como mostra a Figura 20.

3.4.2 Série temporal com dados do *MoCap*

Após a etapa de extração de dados de *MoCap* referentes à marcha e utilizando o vídeo como referência temporal, foi dado prosseguimento à fase de cálculo das coordenadas do CoM para cada *frame* do vídeo de forma automática, a partir da rotina implementada, a uma taxa de amostragem de 120 Hz. A essa taxa, o vídeo que é de aproximadamente 1.3 segundos, teve como saída as coordenadas do CoM para cerca de 240 *frames*.

É importante ressaltar que foi primordial identificar a coluna de dados que representava efetivamente o deslocamento vertical do CoM, ou seja, para que não houvesse

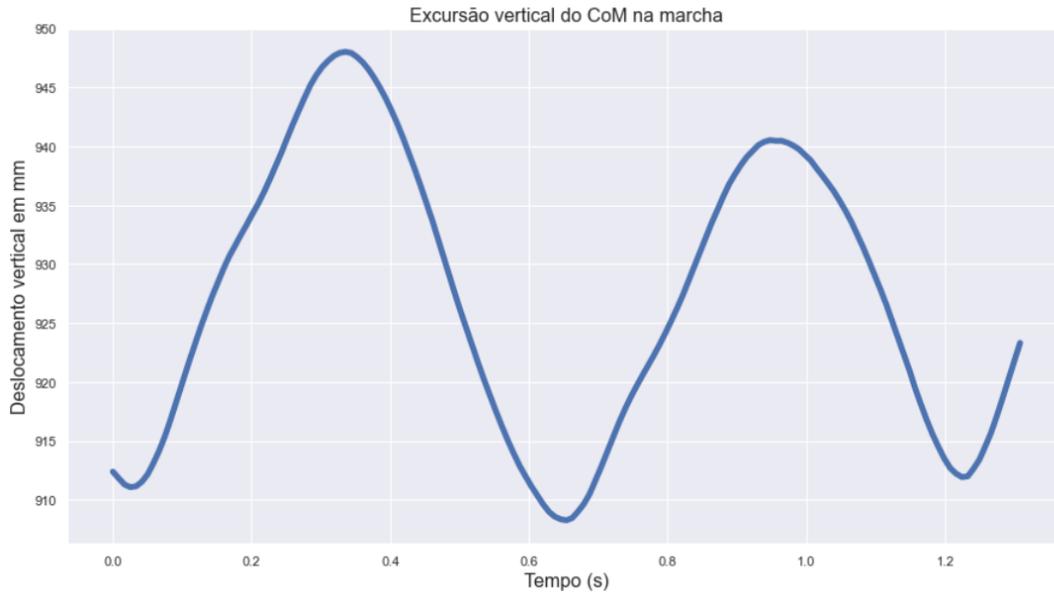


Figura 21 – Excursão vertical do CoM na marcha com dados do *MoCap*

incerteza no momento de visualizar o *plot* da trajetória no tempo. Os eixos x , y e z tratam-se apenas de uma nomenclatura definida no sistema de coordenadas cartesiano e não necessariamente representam os eixos horizontal, vertical e de profundidade, respectivamente.

Sabendo disso, foi observado como se deu a distribuição das variáveis em seus eixos. É possível notar, a partir da Figura 21 que o eixo vertical correspondente à altura pôde ser convencionado como eixo z , uma vez que a trajetória descreve o padrão senoidal na vertical e durante um único ciclo da marcha a posição vertical do CoM possui dois pontos mínimo e máximo, que correspondem aos instantes de contato inicial e apoio médio de cada perna, como comentado em 2.2.

3.5 Excursão vertical do CoM a partir de dados do *BlazePose*

É possível dividir em dois cenários o processo de estimação da trajetória vertical do CoM para a marcha utilizando o *BlazePose*. Esses cenários referem-se às unidades de medidas das coordenadas dos 33 marcadores *3D* espalhados ao longo do corpo, contidos em duas listas de coordenadas que são acessadas através do algoritmo de detecção de pose (seção 3.2).

A primeira é a lista *pose_landmarks*, onde cada marcador consiste no seguinte:

- x e y : Coordenadas dos marcadores em *pixels* normalizadas para $[0, 1]$ pela largura e altura da imagem, respectivamente.
- z : Representa a profundidade do ponto de referência com a profundidade no ponto

médio dos quadris sendo a origem, e quanto menor o valor, mais próximo o ponto de referência está da câmera. A magnitude de z usa aproximadamente a mesma escala de x .

- *visibility*: Um valor entre $[0, 1]$ indicando a probabilidade do ponto de referência ser visível (presente e não ocluído) na imagem.

Portanto, sabendo que o tamanho da imagem é de 800×600 *pixels*, o valor da localização x, y, z de cada marcador e, conseqüentemente, do CoM é dividido pelo tamanho máximo de cada eixo, assim colocando as coordenadas em uma faixa de valores entre 0 e 1.

A segunda é a lista *pose_world_landmarks* que consiste no seguinte:

- x, y e z : Coordenadas *3D* do mundo real (do inglês *real-world coordinates*) ou coordenadas globais, dadas em metros e com a origem fixada no ponto médio entre os quadris.
- *visibility*: Idêntico ao definido no *pose_landmarks* correspondente.

Sendo assim, as coordenadas globais são coordenadas *3D* do mundo real em metros com a origem no centro dos quadris. A lista *WORLD_LANDMARKS* compartilha a mesma topologia que *LANDMARKS*. No entanto, *LANDMARKS* fornece coordenadas (em pixels) de um objeto *3D* projetado na superfície da imagem *2D*, enquanto *WORLD_LANDMARKS* fornece coordenadas (em metros) do próprio objeto *3D*. Este último cenário, em termos de validação dessa ferramenta na apresentação de resultados para a trajetória vertical do CoM na marcha (tema que será discutido nas seções seguintes), seria mais favorável no processo de comparação com o sistema *Mocap*, uma vez que estariam na mesma escala, sendo necessária apenas uma conversão de metros para milímetros no caso das saídas no *BlazePose*.

3.5.1 Série temporal com dados do *BlazePose*

A sequência de imagens exibidas na Figura 22 apresentam 3 *frames* do vídeo do voluntário 15, o primeiro numa posição estática no centro da imagem e os últimos dois durante o ciclo da marcha, acompanhados dos 33 marcadores que formam o modelo de pose humana do *BlazePose*. Nota-se também, assim como na Figura 17, o centro de massa corporal representado pela pequena bolinha amarela localizado na região sacral, um pouco acima dos quadris.

Considerando o mesmo trecho do vídeo em que o sujeito executa o ciclo da marcha, mas com a taxa de amostragem de 30 *Hz* da câmera de visão computacional utilizada no

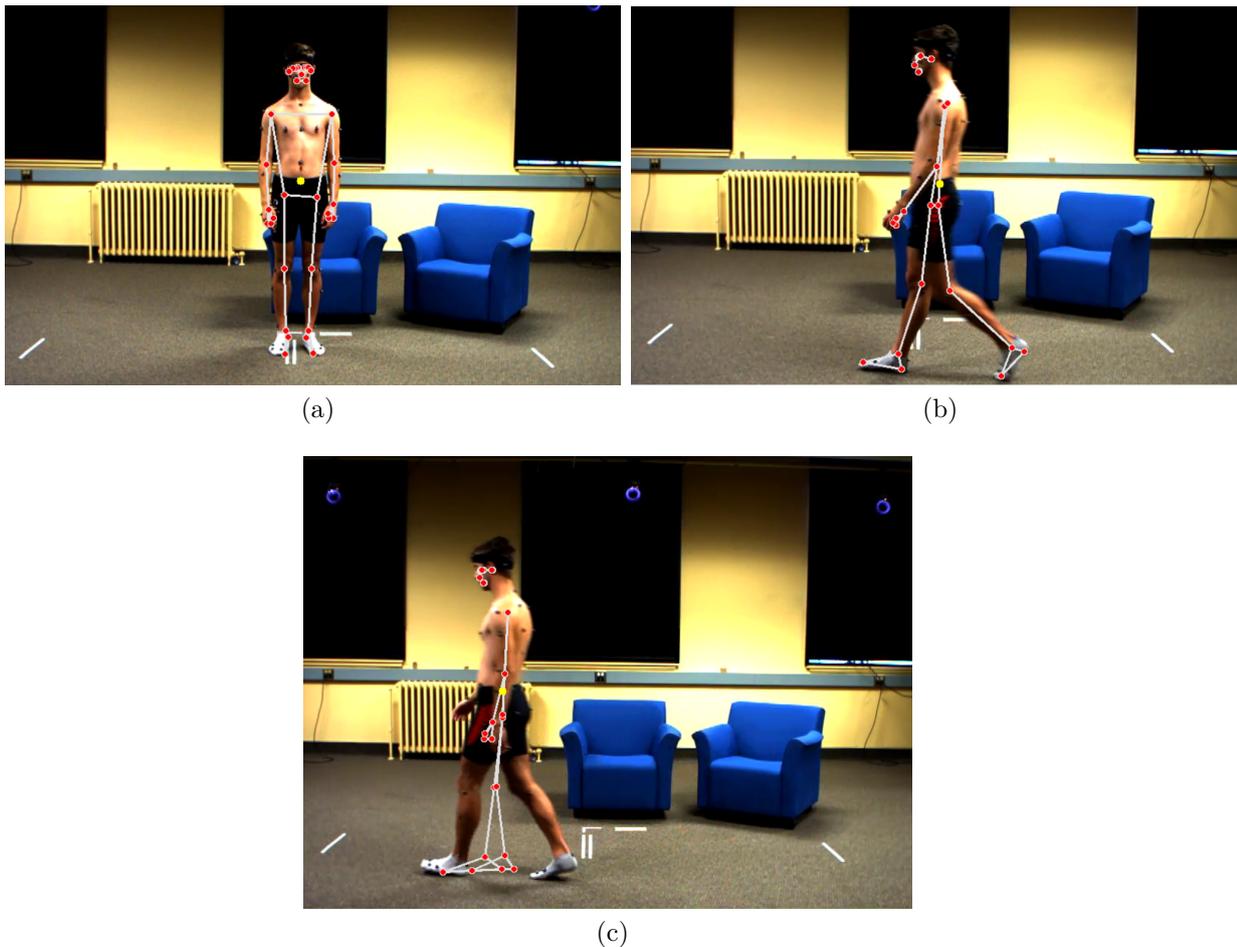
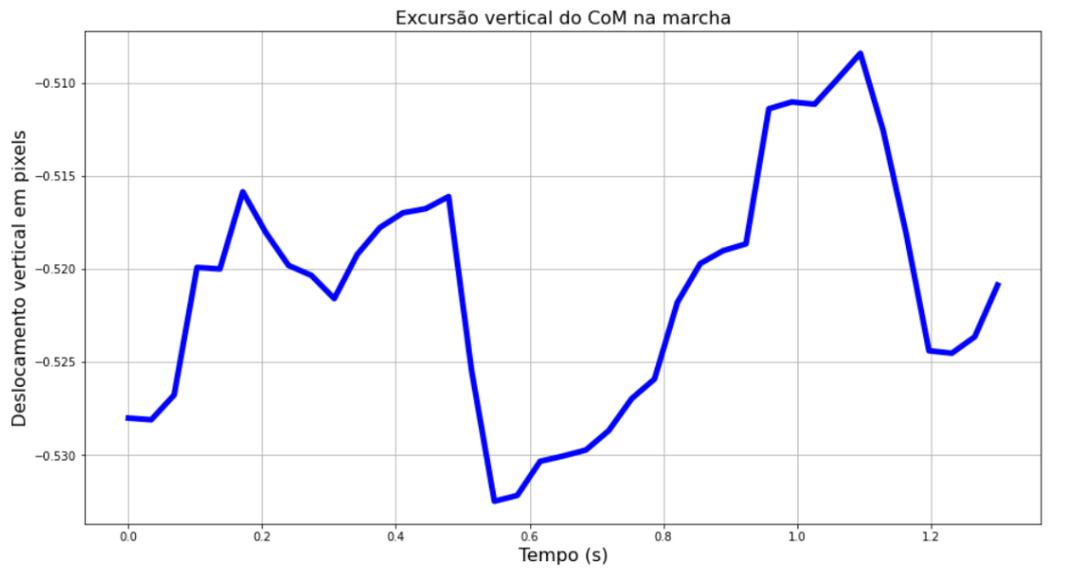


Figura 22 – Centro de massa corporal na posição a) estática; b) início do ciclo da marcha; c) final do ciclo da marcha. O CoM é representado pela bolinha de cor amarela.

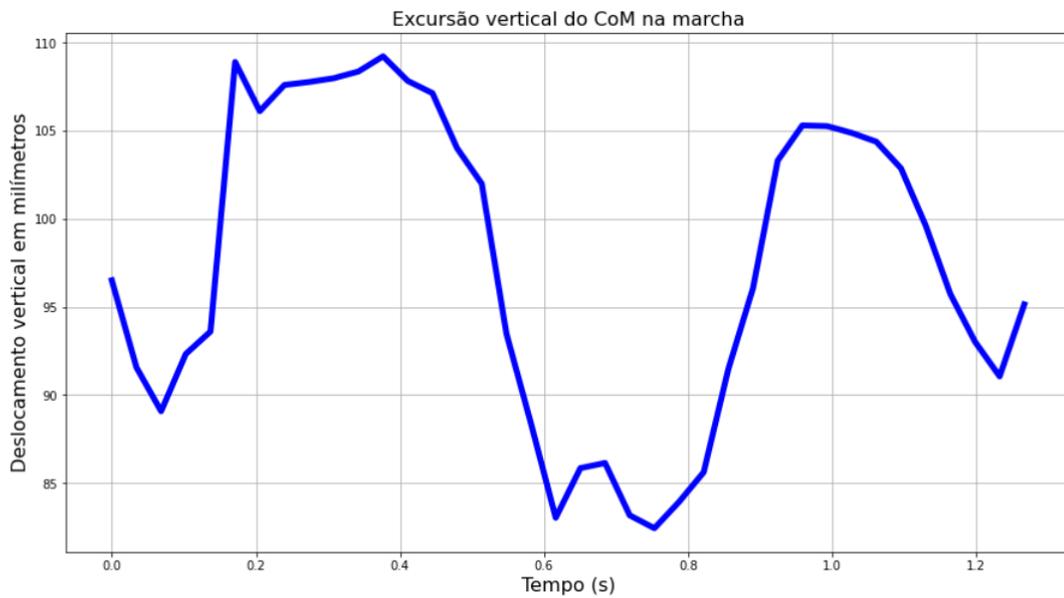
estúdio, a excursão vertical do CoM para esse movimento obtida em *pixels* normalizados, portanto utilizando como parâmetro de escala o *pose_landmarks*, está exposta na letra *a* da Figura 23.

Em um primeiro olhar, pode-se verificar uma tendência de subida e descida característica da forma senoidal da trajetória vertical do CoM porém acompanhada de um conjunto de imperfeições se comparado à série temporal obtida a partir do sistema de captura de movimento cujo resultado foi exibido na Figura 21. A partir desses resultados preliminares, pode-se fazer algumas ponderações:

- Como visto de maneira clara nas imagens da Figura 22, algumas articulações estão ocluídas na imagem e este erro pode afetar a precisão da estimativa do CoM corporal.
- Um fator que pode provocar oscilações indesejadas na trajetória é a forma que os resultados das redes neurais são fornecidos, como esclarecido nos trabalhos de (CAO et al., 2018) e (BAZAREVSKY, 2019). Como os *heatmaps* de saída apresentam uma



(a)



(b)

Figura 23 – Excursão vertical do CoM para a marcha obtida em a) *pixels* normalizados e b) milímetros.

dimensão menor do que a imagem de entrada, sua classificação funciona apenas para uma região da imagem e não para um *pixel*. Por esse motivo, a sobreamostragem do ponto identificado tende a variar dentro da região classificada, provocando as oscilações.

- A origem do sistema com as coordenadas em *pixels* não está localizada no ponto médio dos quadris mas no ponto inferior esquerdo da imagem.
- As coordenadas precisam estar na mesma unidade de medida que as do *MoCap* a fim de que haja uma comparação entre os sistemas.
- É preciso a aplicação de um filtro que resulte em uma trajetória mais estável no tempo.

Examinando a excursão vertical do CoM já convertida para milímetros, portanto utilizando como parâmetro de escala o *pose_world_landmarks*, alcançou-se o seguinte gráfico da letra *b* da Figura 23.

Novamente, observam-se trajetórias com aspectos propensos à forma senoidal porém com muitos ruídos e novos comportamentos. A razão que se dá a essa diferença no gráfico para o observado na Figura 23, é devido à nova posição da origem do sistema, que com a lista *pose_world_landmarks*, se encontra no centro dos quadris. Portanto, a cada *frame* do vídeo a origem se desloca junto com o sujeito, resultando em valores tão baixos para o deslocamento no plano sagital. Desse modo, algumas considerações adicionais devem ser feitas:

- A origem do sistema é móvel e está localizada no ponto médio dos quadris, provocando alterações na excursão vertical.
- As coordenadas estão na mesma unidade de medida que as do *MoCap*, no entanto os valores numéricos da amplitude da excursão vertical do CoM são ainda muito distintos. Logo, convém a representação simultânea da posição e da orientação do sistema de coordenadas do *BlazePose* em relação ao do *MoCap*, através de um processo muito utilizado na robótica conhecido como transformação de coordenadas por matrizes de transformação homogêneas.
- Segue necessária a aplicação de um filtro que resulte em uma suavização do gráfico.

3.5.2 Transformação de coordenadas

Para a transformação de coordenadas, foi utilizado o conceito de construção de matrizes de transformação homogêneas, visto no livro de (SPONG et al., 2006), que podem ser usadas para representar simultaneamente a posição e a orientação de um sistema de

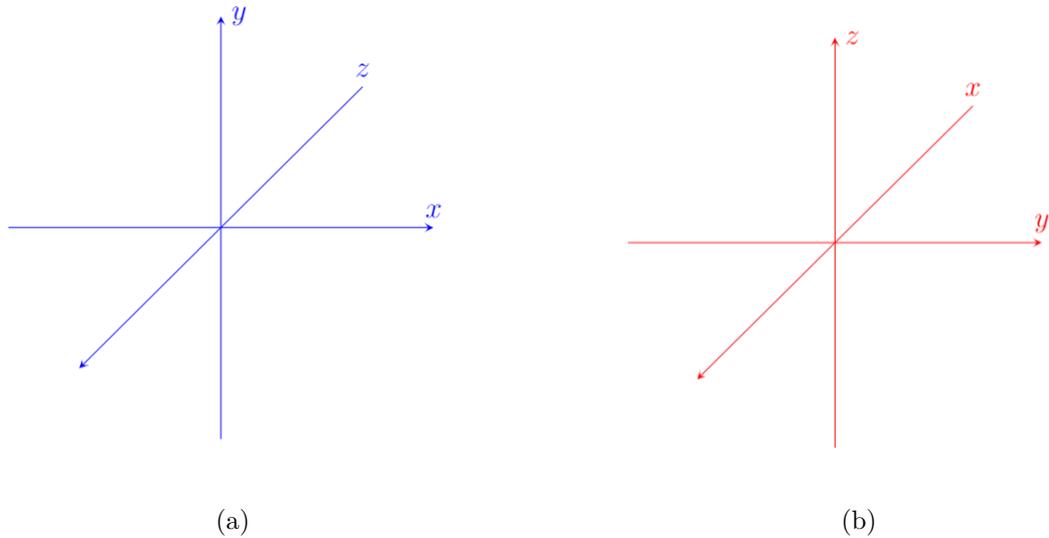


Figura 24 – Orientação dos eixos no sistema do a) BlazePose; b) MoCap

coordenadas em relação a outro. Essas transformações permitiram a representação dos valores da excursão vertical do CoM obtidos do sistema de coordenadas do *BlazePose* no sistema de coordenadas do *MoCap*.

Antes da aplicação desse método, foi necessário que se estabelecesse a orientação dos eixos das coordenadas de cada sistema. Para o *BlazePose*, o próprio sistema já convencionou a nomenclatura dos eixos indicando na saída do algoritmo base, fornecido no material dos desenvolvedores, que o eixo x refere-se ao deslocamento na horizontal, o y na vertical e o z na profundidade. Conseqüentemente, para a definição dos semi-eixos positivo e negativo, bastou verificar, para o sujeito na posição estática no centro da imagem, visto na Figura 22, o valor da posição de marcadores localizados acima e abaixo da origem e também em lados opostos, uma vez que a origem está fixada no ponto médio dos quadris.

Já no *MoCap*, como a origem não está localizada no centro dos quadris, a análise do comportamento dos dados das coordenadas fornecidos foi realizado a partir da própria trajetória do CoM durante os movimentos que o sujeito fazia. A orientação dos eixos para os dois sistemas é, portanto mostrada na Figura 24.

Uma vez que a origem de um sistema de coordenadas é um ponto no espaço, pode-se atribuir coordenadas que representam a posição da origem de um sistema de coordenadas em relação a outro. (SPONG et al., 2006). Dessa forma, é necessário o alinhamento dos dois sistemas através de suas origens, que no *BlazePose* já é conhecida e está localizada no ponto médio dos quadris do sujeito, com as coordenadas em milímetros $x,y,z = (0,0,0)$. No *MoCap*, o cálculo do ponto médio dos quadris do indivíduo se deu através do segmento do vídeo em que ele encontra-se na posição estática de descanso entre um movimento e outro no centro da sala, cujas coordenadas obtidas em milímetros foram $x,y,z = (280, 54,$

928).

O próximo passo foi a aplicação da transformação homogênea ou matriz homogênea. Para o sistema de coordenadas do *BlazePose* estar com seus eixos completamente alinhados ao *MoCap*, 5 transformações foram essenciais, sendo 2 rotações e 3 translações, descritas a seguir:

- Rotação de ângulo $\alpha = 90^\circ$ em torno do eixo y ;
- Seguido por uma rotação de ângulo $\beta = 90^\circ$ em torno do eixo x ;
- Seguido por uma translação de $a = 280mm$ ao longo do eixo x ;
- Seguido por uma translação de $b = 54mm$ ao longo do eixo y ;
- Seguido por uma translação de $c = 928mm$ ao longo do eixo z ;

Vale ressaltar que quando uma transformação é aplicada com relação a um sistema fixo ou de referência, no caso o *MoCap*, a interpretação é exatamente a mesma em relação à composição e ordenação das transformações vistas nas matrizes de rotação, ou seja, foi realizada uma pré-multiplicação da matriz de translação pela matriz de rotação. Portanto, a matriz homogênea seguiu a expressão 3.1.

$$H_B^M = Trans_{x,a} Trans_{y,b} Trans_{z,c} Rot_{y,90^\circ} Rot_{x,90^\circ}. \quad (3.1)$$

A matriz de rotação é representada então pelo produto matricial abaixo,

$$R = \begin{bmatrix} \cos(90^\circ) & 0 & \sin(90^\circ) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(90^\circ) & 0 & \cos(90^\circ) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(90^\circ) & -\sin(90^\circ) & 0 \\ 0 & \sin(90^\circ) & \cos(90^\circ) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

resultando em

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Já a matriz de translação é representada pelo produto matricial abaixo,

$$d = \begin{bmatrix} 1 & 0 & 0 & 280 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 54 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 928 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

que resultou em:

$$d = \begin{bmatrix} 1 & 0 & 0 & 280 \\ 0 & 1 & 0 & 54 \\ 0 & 0 & 1 & 928 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Finalmente a matriz de transformação homogênea será constituída pelo produto matricial entre a matriz de rotação pré-multiplicada pela matriz de translação, mostrada abaixo:

$$H_B^M = \begin{bmatrix} 0 & 1 & 0 & 280 \\ 0 & 0 & -1 & 54 \\ -1 & 0 & 0 & 928 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Com isso, tornou-se possível representar todos os valores do CoM na trajetória vertical obtidos no *BlazePose* no mesmo sistema de coordenadas que o *MoCap*, através da expressão 3.2.

$$P_M = H_B^M \cdot P_B \quad (3.2)$$

$$\begin{bmatrix} P_{Mx} \\ P_{My} \\ P_{Mz} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 280 \\ 0 & 0 & -1 & 54 \\ -1 & 0 & 0 & 928 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} P_{Bx} \\ P_{By} \\ p_{Bz} \\ 1 \end{bmatrix},$$

em que H_B^M é a matriz homogênea, P_B são os valores das coordenadas do CoM no *BlazePose* e P_M são os almejados valores dessas mesmas coordenadas relativos ao sistema de referência *MoCap*. Como o objetivo do trabalho é a avaliação da excursão vertical do CoM, os valores P_{Bx} e P_{By} podem ser desconsiderados, o que resultou apenas nos valores P_{Mz} para representar a trajetória do CoM vertical na marcha.

4 Resultados e Discussões

4.1 Comparação entre os sistemas

A seguinte etapa compreende o processo de transformação de coordenadas e o de filtragem dos sinais com a finalidade de fazer a avaliação da ferramenta *BlazePose* como apropriada ou não para análise de excursão vertical do CoM na marcha, a partir da comparação com os valores dos sinais tomados como referência, que são os obtidos através do sistema *MoCap*.

O resultado para a excursão vertical do CoM após a transformação homogênea é visto na Figura 25. Nota-se que os valores do CoM vertical estão bem mais próximos do que foi visto no *MoCap* (Figura 21), ambos com amplitude em torno de 45 mm. Para atenuar os efeitos de dados ocluídos e outros tipos de ruídos e oscilações indesejadas no gráfico, a próxima etapa foi a de filtragem dos dados.

4.2 Aplicação do filtro *Butterworth*

O ruído nos dados de deslocamento biomecânico costuma ser reduzido por rotinas de filtro digital. Na literatura, interpolações com validação cruzada (WOLTRING, 1986) e o filtro passa-baixas *Butterworth* são frequentemente usados. Segundo o estudo de (ROITHNER et al., 2000), os parâmetros de filtro ideais (frequência e ordem) de um

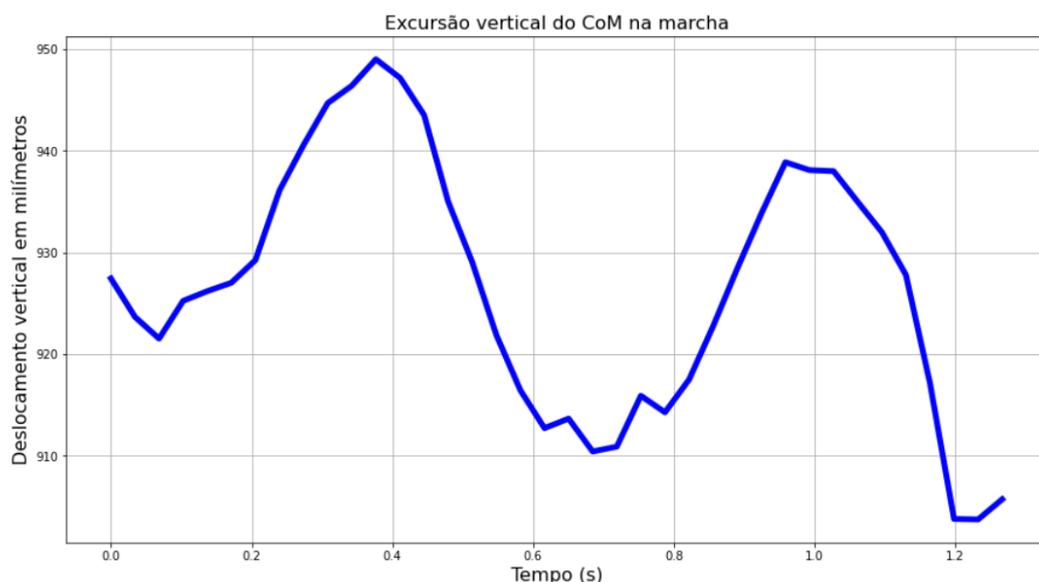


Figura 25 – Excursão vertical do CoM para a marcha obtida em milímetros após transformação homogênea

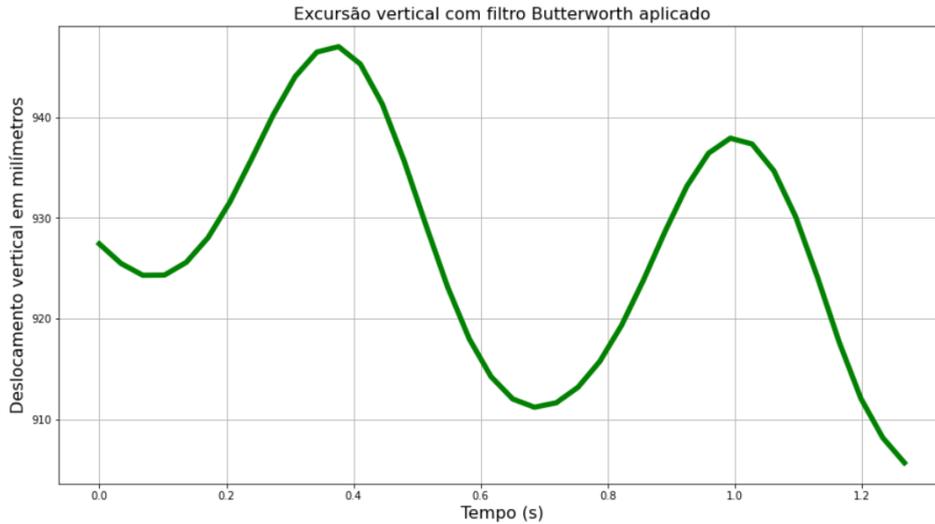


Figura 26 – Excursão vertical do CoM para a marcha obtida em milímetros após transformação homogênea e filtragem

filtro passa-baixas *Butterworth* usando algoritmo de otimização de filtros aplicado a um conjunto de dados cinemáticos na marcha, resulta em um filtro de segunda ordem com frequência de corte representada pela fração da frequência de *Nyquist* usada pelo filtro.

Tendo em vista que o teorema de *Nyquist* enuncia que a frequência de amostragem deve ser pelo menos 2 vezes a frequência máxima do sinal, então, como a frequência de amostragem dos dados do vídeo no *BlazePose* são de 30 Hz, a maior frequência do sinal representado é de 15 Hz, que representa 0.15 vezes a frequência de *Nyquist* ou um valor na faixa de 2 Hz. Dessa forma, as frequências que verdadeiramente representam o sinal, são as frequências mais baixas, portanto, para a filtragem dos dados foi utilizado um filtro *Butterworth* passa-baixas de segunda ordem, em razão de ser um filtro que possui uma resposta razoavelmente rápida e de fácil implementação, reduzindo o custo da aplicação. O resultado da trajetória filtrada pode ser visto na Figura 26, com os efeitos dos ruídos e oscilações indesejadas na excursão vertical do CoM suavizados.

4.3 Aplicação do *Dynamic Time Warping*

As sequências temporais extraídas dos dois sistemas são então comparadas pelo algoritmo de comparação DTW, cujo propósito é identificar as similaridades entre as séries de referência e alvo, como mostra a Figura 27. As entradas reúnem, portanto, tanto os dados da trajetória vertical do CoM na marcha provenientes do sistema de *MoCap*, isto é, os dados de referência, como do alvo, a trajetória a partir da estimativa realizada através do *BlazePose*.

O algoritmo tem início com o cálculo das distâncias euclidianas encontradas entre os dados de referência e o alvo para o movimento selecionado, sendo o cálculo feito a partir

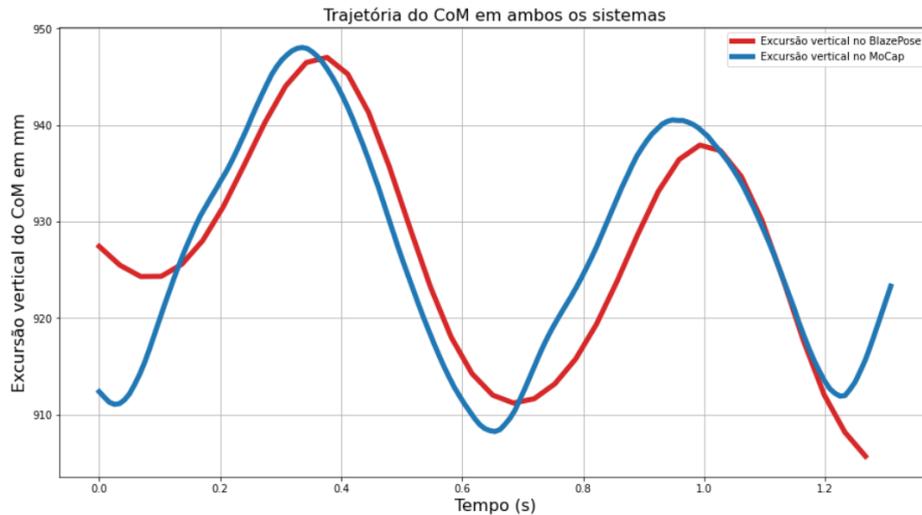
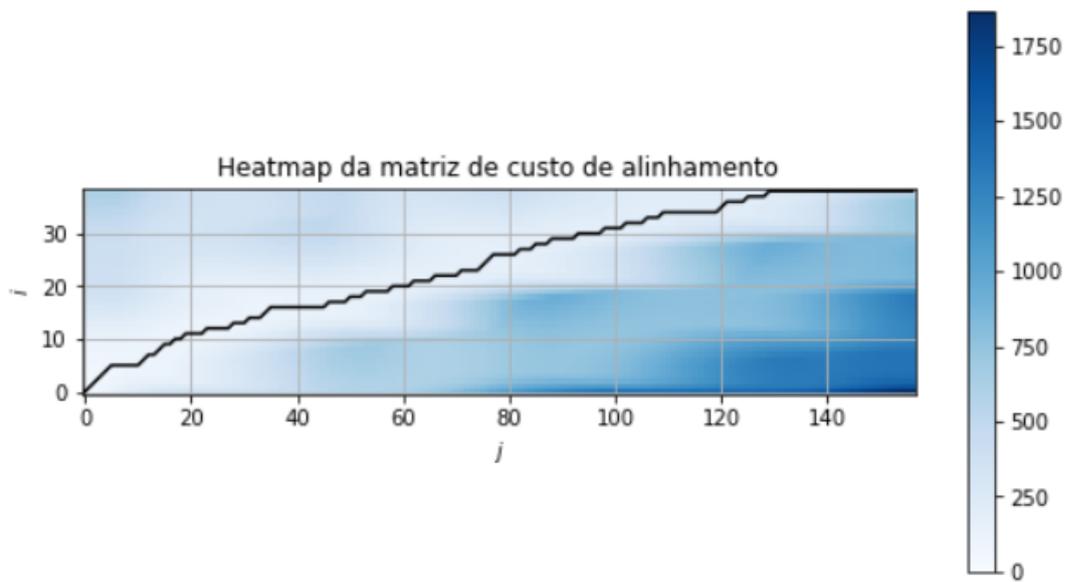


Figura 27 – Excursão vertical do CoM nos dois sistemas para a marcha obtida em milímetros.

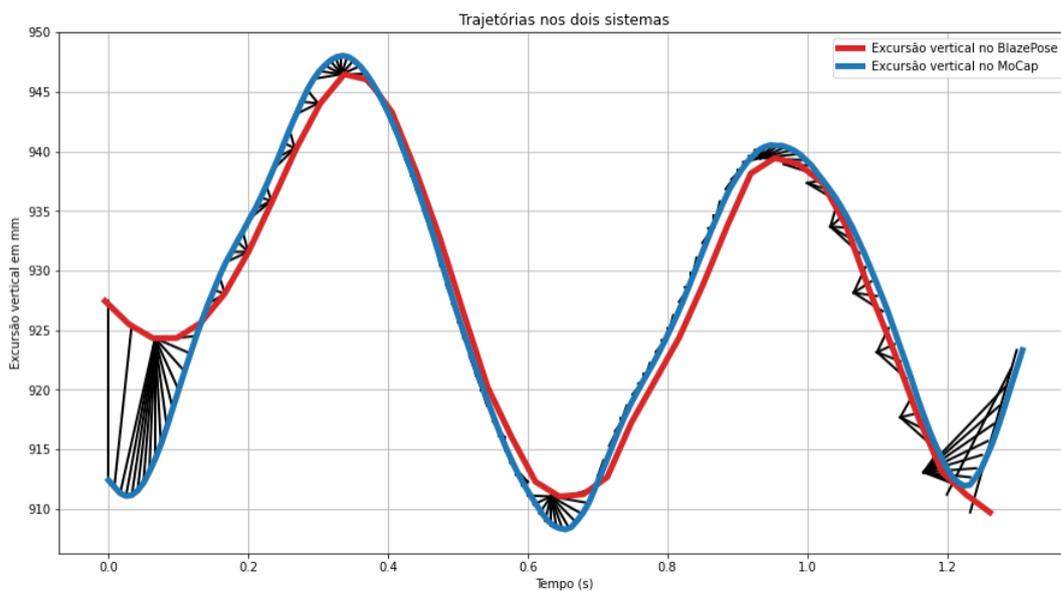
da equação 2.14. As distâncias encontradas entre os sinais são armazenadas em uma matriz chamada matriz de distância euclidiana ou o caminho que configura o mapeamento entre as séries, o *Warping Path*. Partindo dessa matriz, inicia-se o cálculo de custo cumulativo desde o primeiro elemento do canto inferior esquerdo da matriz, adotando-se a equação 2.15, onde são comparados os valores de seus elementos vizinhos, de forma a encontrar o menor custo, ou o menor valor entre eles, e assim seguindo o caminho até o último elemento da matriz.

Deste modo, a matriz pode ser percorrida horizontalmente, verticalmente ou pela diagonal, de acordo com o resultado da comparação dos elementos adjacentes com o elemento atual. Com isso, a matriz é percorrida em direção ao último elemento, localizado no canto superior direito da matriz. O caminho percorrido até esse elemento é armazenado como o caminho que estabelece as correspondências entre os índices de cada uma das séries, a de referência no eixo x e o alvo no eixo y .

A Figura 28.a apresenta o custo cumulativo, em um *heatmap*, onde as maiores distâncias são representadas pelas tonalidades mais escuras da cor azul e o caminho com o menor custo por tonalidades mais claras. O resultado é uma linha irregular em formato de escada, formada a partir das correspondências baseadas nas menores distâncias entre as trajetórias. A Figura 28.b apresenta o alinhamento utilizando o DTW e é possível observar que um único ponto na trajetória do CoM através do *BlazePose* corresponde a vários pontos na trajetória com dados do *MoCap*, o que não necessariamente caracteriza um problema, visto que (KEOGH E. J.; PAZZANI, 2001) afirma que esse comportamento revela um ajuste do algoritmo para a distorção causada no tempo. Pode-se observar que se trata de um alinhamento coerente mesmo com a diferença nas frequências de amostragem dos sinais, que pode ter causado alguns pontos que não coincidem com as curvas em um



(a)



(b)

Figura 28 – a) *Heatmap* com o melhor caminho entre as duas séries pelo DTW; b) Gráfico do DTW aplicado à marcha nos dois sistemas.

trecho do sinal.

Em uma tentativa recente de alinhar dois sinais com frequências de amostragem distintas utilizando DTW, o trabalho de (YIHANG et al., 2020) observa que na prática, o DTW tem um desempenho insatisfatório quando aplicado a sinais biomédicos medidos simultaneamente em diferentes frequências e/ou quando cada sinal individual tem uma frequência de amostragem dinâmica e desigual. Segundo os autores, isso se deve ao problema de singularidade, que resulta de combinação imprecisa de uma única amostra em uma sequência de tempo com um grande número de amostras consecutivas em outro,

resultando em uma distribuição desigual de alinhamentos ponto a ponto.

Tendo em vista o resultado compreensível porém ainda inconclusivo naquele momento, recorreu-se a uma alternativa que utiliza a análise de Fourier para representar as trajetórias em termos dos componentes da série discreta de Fourier ou DFS.

4.4 Análise espectral da trajetória do CoM

Inicialmente, utilizando a equação 2.18, as trajetórias do CoM nos sistemas foram expandidas em termos dos coeficientes a_k da série discreta de Fourier através da aplicação de uma FFT (*Fast Fourier Transform*) em *Python*. O número dos coeficientes DFS calculados para cada coordenada da trajetória vertical do CoM está relacionado a uma energia no domínio da frequência, também calculada em termos dos coeficientes DFS por meio da expressão 2.20.

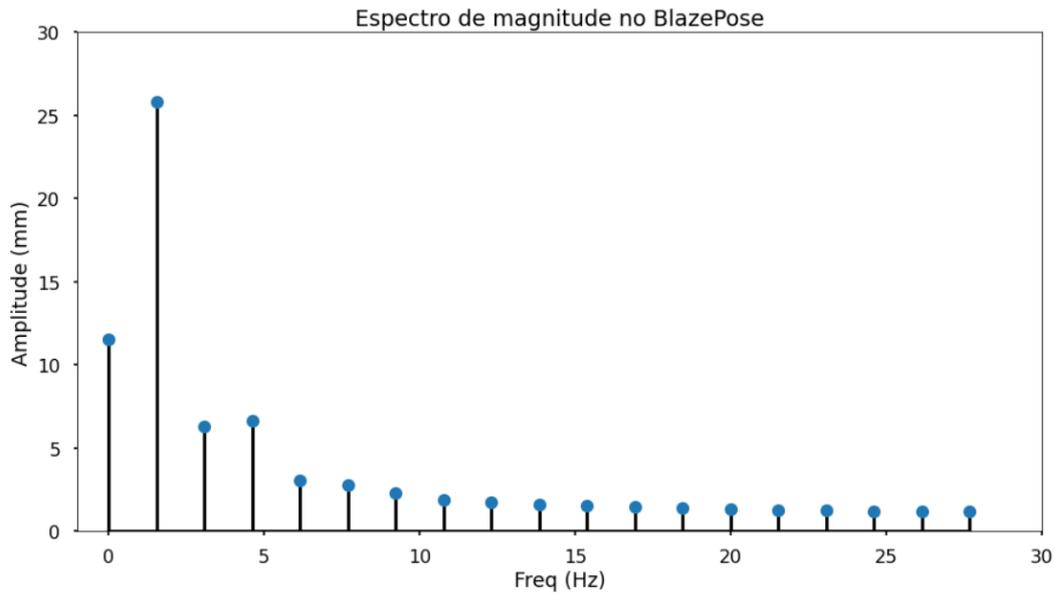
A Figura 29 apresenta o espectro de magnitude do DFS da excursão vertical do CoM no sistema *MoCap* e no *BlazePose*. Em relação ao *MoCap*, a trajetória no domínio do tempo apresenta duas oscilações com amplitudes semelhantes e duração de um ciclo da marcha. O espectro de magnitude do sujeito para este sistema revela dois coeficientes principais para $k = 1$ e $k = 2$. O primeiro constitui aproximadamente 3% da energia total do sinal, enquanto o componente em $k = 2$ tem uma energia de 92%. Este último oscila na frequência 1.53 Hz, muito próxima ao inverso do período do ciclo, em $t = 0.65s$, como visto na Tabela 5.

Tabela 5 – Frequência e período do coeficiente de maior energia do sinal da trajetória vertical do CoM

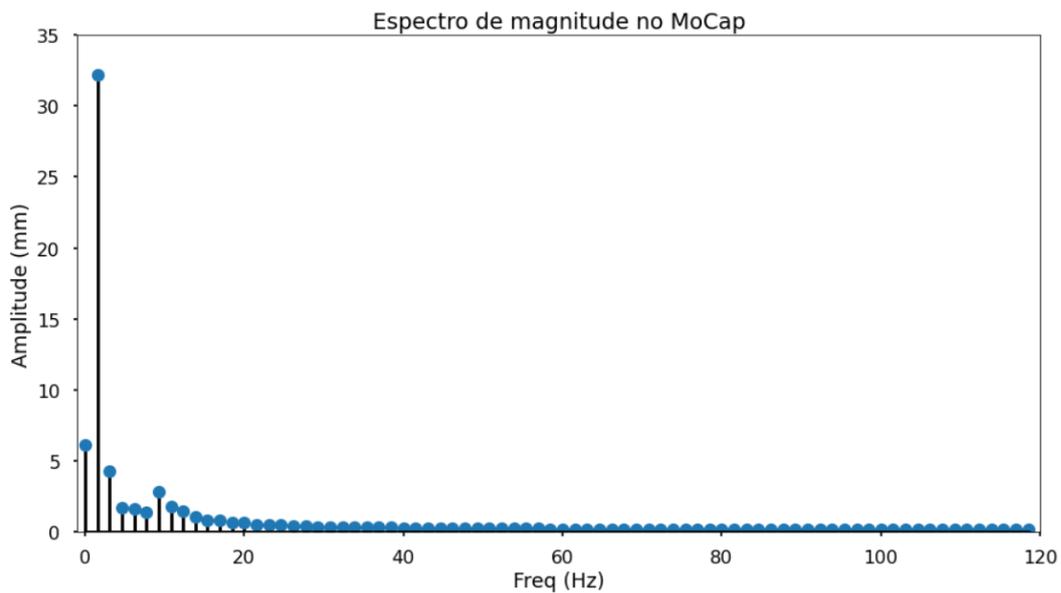
	Frequência do coeficiente (Hz)	Período do coeficiente (s)
<i>MoCap</i>	$f_1 = 1.538$	$t_1 = 0.65$
<i>BlazePose</i>	$f_1 = 1.537$	$t_1 = 0.65$

Já em relação ao *BlazePose*, a energia do sinal é concentrado em dois coeficientes DFS em $k = 1$ e $k = 2$, que possuem aproximadamente 72% e 14% da energia total do sinal. O coeficiente de maior energia oscila na frequência 1.53 Hz e no período $t = 0.65s$, revelando o mesmo comportamento visto no sistema *MoCap*.

É possível observar a partir da Figura 29 e Tabela 5 que embora a contribuição em termos de energia do sinal da trajetória do CoM no *BlazePose* esteja um pouco abaixo do que o visto no *MoCap*, o que é justificado pela estimativa fornecida pela ferramenta, as frequências fundamentais e os períodos coincidem, o que permite o entendimento da análise espectral do CoM como uma alternativa válida para realizar a comparação da similaridade entre as trajetórias verticais nos dois sistemas.



(a)



(b)

Figura 29 – Espectro de magnitude da trajetória vertical do CoM nos sistemas a) *BlazePose* e b) *MoCap*

5 Conclusão

Neste trabalho foi implementada em *Python* uma arquitetura de rede neural convolucional chamada *BlazePose* no cálculo da estimativa da trajetória vertical do centro de massa corporal na marcha, utilizando o método da segmentação. Com o intuito de validar esta ferramenta como uma abordagem alternativa para realizar a estimativa da trajetória do CoM a partir de vídeos, foram utilizadas coordenadas de posição corporal fornecidas por um sistema de captura de movimento óptico (*MoCap*) sincronizadas com um vídeo proveniente da base de dados *MoVi*, que utiliza o sistema *Qualisys* como padrão ouro.

O trabalho mostrou que a ferramenta *BlazePose*, apesar de nova e ainda em processo de aperfeiçoamento pelos desenvolvedores, é capaz de oferecer uma estimativa razoável de pose corporal e cujos dados de posição podem ser utilizados para a estimativa de parâmetros biomecânicos como o centro de massa. Os primeiros resultados experimentais ilustraram, ainda que apenas visualmente, a trajetória do centro de massa em um vídeo demonstrando movimentos simples com uma câmera convencional.

Os resultados experimentais das etapas seguintes, referentes às estimativas das trajetórias do CoM e comparação com as estimativas do sistema de referência (*MoCap*) para a marcha mostraram, no entanto, que não se trata de uma ferramenta ideal, uma vez que oclusões de articulações na imagem, ruídos nos gráficos das trajetórias do CoM e unidades de medidas diferentes do sistema de referência revelaram a necessidade de ajustes do algoritmo com filtragens e transformações de coordenadas.

Por fim, foi realizado o processo de comparação dos sistemas, com o objetivo de alcançar um índice quantitativo que indicasse o quão similares são as trajetórias verticais do CoM em ambos. O fato do CoM apresentar um comportamento periódico fundamentou o uso de uma técnica matemática consolidada na literatura que facilitou a interpretação da trajetória do CoM no domínio da frequência, a análise do espectro de magnitude do CoM. Como esperado, o resultado desta técnica mostrou a similaridade entre as trajetórias por meio dos valores alcançados da frequência fundamental e o seu inverso, o período.

Tendo em vista as limitações ainda existentes no trabalho, algumas etapas podem ser desenvolvidas em trabalhos futuros a fim de consolidar esta proposta como um método alternativo para estimar este parâmetro. Entre as restrições está a análise feita de modo *offline*, devido às várias etapas acumuladas de extração e tratamento de dados. A inclusão de uma análise em tempo real pode ser, portanto, incluída no próximos passos. Outro ponto importante a ser explorado, talvez o mais relevante, é a aprimoração do algoritmo de estimação de pose para lidar de maneira mais robusta com as oclusões e presença de ruídos nas imagens através da construção de filtros, a fim de atenuar estes efeitos na análise

da trajetória do CoM. Ainda, investigar algumas questões que não ficaram claras até o momento como, por exemplo, o algoritmo que os desenvolvedores do *BlazePose* utilizaram para definir os parâmetros de distância entre a câmera que grava o vídeo e as pessoas e objetos contidos no espaço gravado, que geram as coordenadas globais fornecidas.

Referências

- ABRAHAM, A. Meta-learning evolutionary artificial neural networks. *Neurocomputing Journal*, Elsevier Science, v. 56, 2004. Citado na página [28](#).
- ANDRILUKA, M. Posetrack: A benchmark for human pose estimation and tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Elsevier Science, p. 5167—5176, 2018. ISSN 10636919. Citado na página [31](#).
- BAPTISTA, R. D. S. *Automatic Human Movement Assessment with Switching Linear Dynamic System: Motion Segmentation and Motor Performance*. [S.l.: s.n.], 2017. v. 25. 628—640 p. ISSN 15344320. Citado na página [36](#).
- BARTONEK A., S. H. Comparison and evaluation of two common methods to measure center of mass displacement in three dimensions during gait. *ScienceDirect, Human Movement Science*, v. 25, n. 2, p. 238–256, 2006. Citado na página [21](#).
- BAZAREVSKY, V. Blaze-face: Sub-millisecond neural face detection on mobile gpus. *CoRR, abs/1907.05047*, 2019. Citado 3 vezes nas páginas [32](#), [33](#) e [54](#).
- BAZAREVSKY, V. et al. Blazepose: On-device real-time body pose tracking. *arXiv*, Elsevier Science, 2020. Citado 5 vezes nas páginas [11](#), [31](#), [32](#), [33](#) e [41](#).
- BISHOP, C. Neural networks for pattern recognition. *Oxford University Press*, Elsevier Science, 1995. Citado 2 vezes nas páginas [11](#) e [29](#).
- BUTTERWORTH, S. et al. *On the theory of filter amplifiers*. *Wireless Engineer*. [S.l.: s.n.], 1930. v. 7. 536—541 p. Citado 2 vezes nas páginas [35](#) e [36](#).
- CAO, Z. et al. Openpose: realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1812.08008*, Elsevier Science, 2018. Citado 3 vezes nas páginas [11](#), [31](#) e [54](#).
- CAVAGNA, G. et al. Mechanical work in terrestrial locomotion: two basic mechanisms for minimizing energy expenditure. *American Journal of Physiology*, 1977. Citado na página [25](#).
- CHENG S.Y.; TRIVEDI, M. Human posture estimation using voxel data for “smart” airbag systems: Issues and framework. *In Proceedings of the IEEE Intelligent Vehicles Symposium*, 2004. Citado na página [26](#).
- CORAZZA S.;UNDERMANN, L. A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach. *Revista portuguesa de educação, Annals of Biomedical Engineering*, v. 34, n. 6, 2006. Citado na página [27](#).
- DEMPSTER, G. R. L. G. W. T. Properties of body segments based on size and weight. *American Journal of Anatomy*, 1967. Citado 2 vezes nas páginas [21](#) e [22](#).

DIAZ, C. P. O. Characterization of amputee gait using a biomechanical approach. 2015. Citado 6 vezes nas páginas 11, 19, 26, 27, 38 e 39.

DINH, D. et al. Real-time 3d human pose recovery from a single depth image using principal direction analysis. *Appl. Intell.*, 2014. Citado na página 27.

DRILLIS R.; CONTINI, R. Body segment parameters. *New York: Office of Vocational Rehabilitation*, v. 3, n. 1166, 1966. Citado na página 21.

GARD, S. et al. Comparison of kinematic and kinetic methods for computing the vertical motion of the body center of mass during walking. *Human Movement Science Journal*, 2004. Citado 2 vezes nas páginas 22 e 25.

GHORBANI, S. et al. Movi: A large multipurpose motion and video dataset. *Scholars Portal Dataverse*, 2020. Citado 7 vezes nas páginas 12, 13, 46, 47, 48, 49 e 50.

GÜLER, R. et al. Densepose: Dense human pose estimation in the wild. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Elsevier Science, p. 7297—7306, 2018. ISSN 10636919. Citado na página 31.

HORNIK, K. et al. *Multilayer feedforward networks are universal approximators*. [S.l.: s.n.], 1989. v. 2. 359–366 p. ISSN 08936080. Citado na página 30.

JAFFREY, M. E. Estimating centre of mass trajectory and subject-specific body segment parameters using optimisation approaches. *Ph.D. Thesis, Victoria University, Melbourne, Australia*, ScienceDirect, 2008. Citado na página 22.

JIAN Y.; WINTER, D. Trajectory of the body cog and cop during initiation and termination of gait. *Gait and Posture*, v. 1, p. 9–22, 1993. Citado na página 25.

JUNG C.K.; PARK, S. Compliant bipedal model with the center of pressure excursion associated with oscillatory behavior of the center of mass reproduces the human gait dynamics. v. 47, 2014. Citado na página 19.

K., H. et al. Minimal marker set for center of mass estimation in running. *Gait Posture*, ScienceDirect, v. 25, n. 4, p. 552–555, 2009. Citado na página 21.

KAICHI, T. et al. Image-based center of mass estimation of the human body via 3d shape and kinematic structure. *Sports Engineering*, 2019. Citado na página 25.

KEOGH E. J.; PAZZANI, M. J. *Derivative Dynamic Time Warping*. [S.l.: s.n.], 2001. 1—11 p. Citado 5 vezes nas páginas 11, 36, 37, 38 e 63.

LAFOND D., D. M. P. F. Comparison of three methods to estimate the center of mass during balance assessment. *J. Biomech*, ScienceDirect, v. 37, p. 1421–1426, 2004. Citado na página 22.

LOPER, M. et al. Mosh: motion and shape capture from sparse markers. *ACM Transactions on Graphics*, v. 33, p. 1–13, 2014. Citado na página 47.

MACKINNON. Control of whole body balance and posture in the frontal plane during walking. *J. Biomech.*, v. 26, p. 633–644, 1993. Citado na página 25.

- MCKINON, W. et al. The agreement between reaction-board measurements and kinematic estimation of adult male human whole body centre of mass location during running. *Physiological Measurement*, 2004. Citado na página 19.
- MUSCOLO G.; HASHIMOTO, K. A method for the calculation of the effective center of mass of humanoid robots. *2011 11th IEEE-RAS International Conference on Humanoid Robots Bled*, 2011. Citado na página 19.
- PAN S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, IEEE.*, Elsevier Science, p. 1345–1359, 2010. ISSN 10414347. Citado na página 30.
- PATLA F.; WINTER, D. Stiffness control of balance in quiet standing. *J. Neurophysiol.*, 1998. Citado na página 25.
- ROITHNER, R. et al. Determination of optimal filter parameters for filtering kinematic walking data using butterworth low pass filter. *18 International Symposium on Biomechanics in Sports (2000)*, 2000. Citado na página 61.
- SAINI, M. et al. The vertical displacement of the center of mass during walking: A comparison of four measurement methods. *J. Biomech Eng.*, v. 120, 1998. Citado na página 21.
- SAKOE H.; CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING*, 1978. Citado na página 36.
- SHOTTON, J. et al. Real-time human pose recognition in parts from single depth images. *Commun.*, v. 56, p. 116–124, 2013. Citado na página 27.
- SPONG, M. et al. *Robot Modeling and Control*. [S.l.: s.n.], 2006. v. 1. 29–57 p. Citado 6 vezes nas páginas 11, 33, 34, 35, 56 e 57.
- TOSHEV, A. et al. Deeppose: Human pose estimation via deep neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Elsevier Science, p. 1653–1660, 2014. Citado na página 29.
- WINTER, D. *Biomechanics and motor control of human movement*. [S.l.: s.n.], 1990. v. 28. 82–104 p. Citado 6 vezes nas páginas 11, 13, 21, 23, 24 e 26.
- WOLTRING, H. A fortran package for generalized, cross-validatory spline smoothing and differentiation. *Advances in Engineering Software (1978)*, v. 8, p. 104–113, 1986. Citado na página 61.
- WU P.; DIETERRICH, T. et al. Improving svm accuracy by training on auxiliary data sources. *School of EECS, Oregon State University, Corvallis, OR 97331*, Elsevier Science, 2004. Citado na página 30.
- XIU, Y. Pose flow: Efficient online pose tracking. *British Machine Vision Conference 2018, BMVC 2018*, p. 1–12, 2019. Citado na página 31.
- YIHANG, J. et al. Eventdtw: An improved dynamic time warping algorithm for aligning biomedical signals of nonuniform sampling frequencies. *Sensors*, n. 9, 2020. Citado na página 64.

ZATSIORSKY ET AL., . Adjustments to zatsiorsky-seluyanov's segment inertia parameters. *Journal of Biomechanics*, ScienceDirect, v. 29, n. 9, p. 1223–1230, 1996. Citado na página [22](#).

ZHANG X.; GONZÀLEZ, J. G. W. Human pose estimation from monocular images: A comprehensive survey. *Sensors (Switzerland)*, 2016. Citado na página [27](#).